Université Paris VII - Denis Diderot

École Doctorale Paris Centre

Thèse de doctorat

Discipline : Mathématiques Appliquées

présentée par

Régis Lebrun

Contributions à la modélisation de la dépendance stochastique

---

dirigée par Josselin Garnier

Rapporteurs : M. Fabrizio Durante Libera Università di Bolzano
M. Fabrice Gamboa Université Paul Sabatier

Soutenue le 24 mai 2013 devant le jury composé de :

M. Jean-François Delmas	École des Ponts ParisTech	Président
M. Fabrice Gamboa	Université Paul Sabatier	Rapporteur
M. Josselin Garnier	Université Paris Diderot	Directeur
M. Fabien Mangeant	EADS Innovation Works	Examinateur
Mme. Nadia Oudjane	EDF R&D	Examinatrice
Mme. Dominique Picard	Université Paris Diderot	Examinatrice

LPMA / UFR de Mathématiques
Site Chevaleret
75 205 Paris Cedex 13

École doctorale Paris centre Case 7012
Site Chevaleret
75 205 Paris Cedex 13

A mon épouse Anne et mes deux filles Alice et Charlotte,
pour leur amour, leur soutien et leur patience...

Remerciements

Josselin Garnier, dont la gentillesse n'a d'égale que l'excellence mathématique, m'a fait l'honneur d'accepter d'encadrer cette thèse. C'était un pari risqué, le candidat ayant 36 ans révolus, ayant une famille à charge et une activité professionnelle à conduire. Je le remercie chaleureusement pour la confiance qu'il m'a accordée. Sans ses conseils, ses relectures minutieuses et exigeantes, ses questions précises, ce travail n'aurait certainement pas atteint sa maturité actuelle.

Cette thèse n'aurait évidemment jamais pu se faire sans le soutien de mon employeur EADS, qui, en les personnes d'Eric Duceau et de Fabien Mangeant, m'a accordé du temps et des moyens pour mener à bien ce projet. Je les remercie pour la liberté de travail qui m'a été accordée et j'espère que le produit de ce travail sera à la hauteur de l'investissement qui a été fait.

Je remercie les différents membres du jury. C'est un grand honneur qu'ils me font en acceptant cette charge, qui implique également d'avoir eu le courage de se plonger dans la lecture de mes travaux.

Je remercie également François Dubois, sans qui l'idée de préparer cette thèse serait restée une pensée fugace. Il m'a convaincu qu'une telle entreprise avait du sens tant pour moi que pour mon employeur. Il a ouvert le cahier, il ne restait plus qu'à écrire.

Je remercie tous mes collègues d'Innovation Works, d'Airbus, d'Astrium, d'Eurocopter, d'Électricité de France, de Phimeca, de l'ONERA et de tant d'autres endroits, qui, lorsque je leur parlais de mes travaux, ont montré de l'intérêt et m'ont encouragé.

Enfin, Anne Dutfoy, mon épouse, co-auteur de plusieurs des travaux présentés dans cette thèse, m'a soutenu tout au long de cette aventure de quarante mois. Plus que tout autre, elle mérite mes remerciements. Elle m'a réconforté dans les périodes de doute quand je cherchais mes thèmes de travail, a partagé mon enthousiasme quand les difficultés étaient enfin surmontées. Sa foi inaltérable dans le succès de cette entreprise a été mon plus solide soutien. Ce sont infiniment plus que des remerciements que je lui adresse.

Contents

Introduction

The modeling of stochastic dependence is an old subject both in theoretical and applied probability and statistics. From a statistical point of view, the pioneers were Pearson (1857-1936) and Spearman (1863-1945), who introduced the concept of coefficient of correlation to quantify by a scalar value the link between the joint measurement of two quantities.

It is known from the elementary course in probability that the behaviour of a random vector is given by its joint distribution function, and not only by the collection of its marginal distribution functions. But when it comes to model real life random systems, or to assess the robustness of an industrial system, the engineering practices are sometimes quite far away from the theory. The joint distribution is either obtained from the marginal distributions and the strong hypothesis of independent components, or the modelling is restricted to the most classical multivariate distributions (Gaussian vectors, multinomial distribution). To make things clear, we give an introduction to dependence modeling in Chapter 1. There is no original material in this chapter, but it gives the basis to all the other chapters. In particular, it defines the notion of copula, which is exactly the mathematical object that represents fully the dependence structure of a random vector with absolutely continuous marginal distributions, which will be the setting of all the manuscript but the last chapter.

In the field of probabilistic uncertainty management, the objective is to quantify several performance criteria related to a given system based on the ouput $\boldsymbol{Y}$ of a numerical model g of this system and a probabilistic modeling of a set of uncertain parameters of this model, grouped into a random vector $\boldsymbol{X}$, and such that:

$$\boldsymbol{Y} = g(\boldsymbol{X}) \tag{1}$$

Formally, the aim of the probabilistic uncertainty management is to gain knowledge on the joint distribution of $\boldsymbol{Y}$ based on the knowledge of g and the distribution of $\boldsymbol{X}$. Typical performance criteria are the evaluation of probabilities of failure $\mathbb{P}(\boldsymbol{Y} \in B)$ for given events B that characterize the behaviour of a complex system.

A common belief in uncertainty management is that the full knowledge of the exact joint distribution function may not be needed to compute eg. probabilities of failure, as the decision making is mainly done based on the order of magnitude of these probabilities rather than their exact value.

Another belief is that the joint distribution of a set of random variables is out of reach of the practitioner, mainly due to the lack of statistical data. In this case, the probabilistic modeling is either to consider independent random variables (which is a particular case of dependence), or to summarize the dependence into a set of scalars, most of the time a set of linear correlation coefficients.

As a result of these beliefs, any method able to compute a probability of failure based on marginal information and partial dependence information should provide a meaningful

result. But it should question people! From a methodological point of view, what is the point of spending a lot of efforts in the definition of g, which is most of the time a computationaly intensive simulation software based on the numerical integration of several coupled partial differential equations that needed a tremendous amount of development and validation work, if one try to quantify $\boldsymbol{Y}$ based only on a poor modeling of $\boldsymbol{X}$? From a probabilistic point of view, if there exists a procedure able to compute $\mathbb{P}(\boldsymbol{Y} \in B)$ for any Borelian set B and any measurable function g, based only on a partial description of the joint distribution function of $\boldsymbol{X}$, then the missing information is hidden in the procedure! For a given $\boldsymbol{x} \in \mathbb{R}^n$, if we take $B = \{1\}$, $g(\boldsymbol{s}) = \mathbb{1}_{s_1 \leq x_1, \ldots, s_n \leq x_n}$, then the procedure computes $\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$ for any $\boldsymbol{x} \in \mathbb{R}^n$. It is exactly the evaluation of the joint distribution of $\boldsymbol{X}$ at $\boldsymbol{x}$, which was supposed to be unknown!

The first belief must certainly be assessed by at least some experimental studies, which is precisely the aim of the work presented in Chapter 2. It appears that in some situations, it may be possible to get a reasonable estimate of probabilities of failure for any joint distribution function that reproduces the available information on marginal distributions and dependence structure, but it is certainly not a generic situation, in particular it is possible to build generic examples for which the value of a probability of failure varies by several orders of magnitude while preserving both the marginal distributions and the correlation matrix. This work has been published in [DL09].

From a statistical point of view, the second belief is mostly true if one work in high dimension due to the well-known *curse of dimensionality*. But instead of hidding the copula in the computational algorithm, it may be desirable to let the practioner choose it explicitely. Even if there is no statistical evidence to assess it, at least the practitioner knows exactly what he is computing. In this view, the use of the Nataf transformation is a common practice to induce dependence amongst random variables with known marginal distributions and known correlation matrix. The detailed analysis of this transformation allows to make explicit the choice of dependence structure it induces and its consequences on uncertainty quantification such as the evaluation of probabilities of failure. This work is presented in Chapter 3, and has been published in [LD09c]

This analysis of the Nataf transformation allows also to generalize this transformation to a broader class of copulas, leading to the generalized Nataf transformation. This analysis is the occasion to present classical material such as the first and second order reliability methods (FORM and SORM) and the notion of standard space in a slightly unusual way that we think is more adapted to the mathematical analysis. This work is presented in Chapter 4, and has been published in [LD09b].

Often presented as an alternative to the Nataf transformation, the Rosenblatt transformation is another tool to express the probability of failure with respect to an arbitrary distribution function for $\boldsymbol{X}$ as a probability of failure for an event involving a standard multi-dimensional normal distribution. Most of the time, this transformation is mentioned only to say that it needs the full definition of the joint distribution function, which is supposed to be unreachable, and even if one knows this distribution function, the results of the FORM and SORM approximations are dependent on the choice of conditioning order involved in the definition of the Rosenblatt transformation. In Chapter 5, we study in detail both the relationship between the Rosenblatt transformation and the Nataf transformation and the effect of the conditioning order on the Rosenblatt transformation. In particular, we show that both transformations are equal in the case of a joint distribution with normal copula, which is the copula underlying the Nataf transformation as seen in Chapter 3. We also study the possible extension of the Rosenblatt transformation in the way the Nataf transformation has been generalized in Chapter 4. This work has been

published in [LD09a].

In some applications, one has to model the joint distribution of a random vector with given marginal distributions and with a physical constraint that must be satisfied almost surely. A typical situation is when the components (or a strictly monotonic function of the components) of the vector have to be ordered in a given way. This modeling problem is made of two parts: what are the compatibility conditions between the marginal distributions and the constraints, and for compatible marginal distributions, what are the possible copulas? We study in details this situation in Chapter 5, for the case where it is possible to reduce the modeling problem to the determination of the possible marginal distributions and copulas for the joint distribution function of order statistics, a problem that has not yet been addressed in the literature as far as we know. While the compatibility conditions are well-known for the marginal distributions, the compatibility condition for the copulas is new. As it excludes all the classical continuous copulas, we also present a new family of bi-dimensional copulas well suited to this modeling situation. We prove the existence and uniqueness of the copula with the largest possible support with reasonable assumptions, and we give all the needed algorithmic details for an actual use of these copulas in simulation. This work is presented in Chapter 6, and has been proposed for publication in [LD].

The use of copulas to study the dependence structure of random vectors is particulary relevant when the marginal distributions are continuous, as there is a one-to-one relation between copulas and joint distribution functions. For discrete distributions, things are less clear. There is no unique copula associated to a given multivariate discrete distribution, and the transposition of habits gained in the continuous case to the discrete case can lead to erroneous conclusions. As such, there is no clear gain to separate the copula and the marginal distributions for models such as the multinomial distribution, the multi-dimensional hypergeometric or multi-dimensional Pólya distributions. Despite the tremendous amount of literature on these models, mainly due to their wide range of application in many fields of probability and statistics, no efficient algorithm is available to compute their joint distribution function (or more generally rectangular probabilities) for high-dimensional applications. We develop such an algorithm and prove its efficiency both theoretically and numericaly. This algorithm is presented in Chapter 7, and has been published in [Leb12].

We give some insight of possible future work on dependence modeling in the conclusion.

Chapter 1

Introduction to dependence modeling

In this chapter, we recall the main results on dependence modeling that will be used in the remainder of the thesis. The interested reader will find a detailed introdution to this theory in [Nel06], [Joe97] and [ELM03], from which the results of this section have been extracted.

The joint distribution function of a random vector plays a central role in the probabilistic modeling as it fully describes the probability distribution of the phenomenon under study. Such a function is defined by:

Definition 1.1. Let $\mathcal{E} = (\Omega, \mathcal{B}(\Omega), \mathbb{P})$ be a given probability space, $\boldsymbol{X}$ a n-dimensional random vector defined on $\mathcal{E}$ and taking values in $\mathbb{R}^n$, μ_X the push forward probability measure of $\mathbb{P}$ by $\boldsymbol{X}$, i.e such that:

$$\forall B \in \mathcal{B}(\mathbb{R}^n), \quad \mu_X(B) = \mathbb{P}\left(\boldsymbol{X}^{-1}(B)\right)$$

where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-field of $\mathbb{R}^n$. The **joint distribution function** $F_{\boldsymbol{X}}$ of $\boldsymbol{X}$ (or simply its distribution function F if there is no confusion) is the function defined by:

$$F_{\boldsymbol{X}} : \left| \begin{array}{ccl} \mathbb{R}^n & \to & [0,1] \\ \boldsymbol{x} & \mapsto & F_{\boldsymbol{X}}(\boldsymbol{x}) = \mu_{\boldsymbol{X}}((-\infty, x_1] \times \ldots \times (-\infty, \ldots, x_n]) = \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) \end{array} \right.$$

and when $\boldsymbol{X}$ is absolutely continuous, its density function $p_{\boldsymbol{X}}$ is given by:

$$p_{\boldsymbol{X}} : \left| \begin{array}{ccl} \mathbb{R}^n & \to & \mathbb{R}^+ \\ \boldsymbol{x} & \mapsto & p_{\boldsymbol{X}}(\boldsymbol{x}) = \dfrac{\partial^n F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x_1 \ldots \partial x_n} \end{array} \right.$$

and is such that:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} p_{\boldsymbol{X}}(\boldsymbol{\xi}) \, \mathrm{d}\xi_1 \ldots \mathrm{d}\xi_n$$

The next proposition gives some properties of a joint distribution function:

Proposition 1.2. *Let $F_{\boldsymbol{X}}$ be the joint distribution function of a n-dimensional random vector $\boldsymbol{X}$.*

1. *The function $F_{\boldsymbol{X}}$ is n-increasing:*

$$\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \quad \sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1 + \ldots + i_n} F_{\boldsymbol{X}}(\xi_1^{i_1}, \ldots, \xi_n^{i_n}) \geq 0$$

with $\forall j \in \{1, \ldots, n\}$, $\xi_j^1 = a_j$ and $\xi_j^2 = b_j$.

2. *The function $F_{\boldsymbol{X}}$ is right-continuous:*

$$\lim_{x'_1 \downarrow x_1, \ldots, x'_n \downarrow x_n} F_{\boldsymbol{X}}(\boldsymbol{x}') = F_{\boldsymbol{X}}(\boldsymbol{x})$$

and such that

$$\lim_{x_1 \to +\infty, \ldots, x_n \to +\infty} F_{\boldsymbol{X}}(\boldsymbol{X}) = 1$$

3. *$\forall \boldsymbol{x} \in \mathbb{R}^n$, the left limit of $F_{\boldsymbol{X}}$ at $\boldsymbol{x}$ exists and is equal to $\mathbb{P}(X_1 < x_1, \ldots, X_n < x_n)$:*

$$\lim_{x'_1 \uparrow x_1, \ldots, x'_n \uparrow x_n} F_{\boldsymbol{X}}(\boldsymbol{x}') = \mathbb{P}(X_1 < x_1, \ldots, X_n < x_n)$$

The interest of such a function is that it fully characterizes the distribution $\mu_{\boldsymbol{X}}$ of $\boldsymbol{X}$:

Theorem 1.3. *The joint distribution function $F_{\boldsymbol{X}}$ of a n-dimensional random vector $\boldsymbol{X}$ characterizes its distribution $\mu_{\boldsymbol{X}}$: two n-dimensional random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution $\mu_{\boldsymbol{X}} = \mu_{\boldsymbol{Y}} = \mu$ if and only if they have the same joint distribution function $F_{\boldsymbol{X}} = F_{\boldsymbol{Y}} = F$. The probability measure of $\boldsymbol{X}$ can then be denoted either by $\mu_{\boldsymbol{X}}$ or by μ_F.*

Proof. By the definition of a joint distribution function, if the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ share the same probability measure, they have the same joint distribution function.

On the other side, the collection $\mathcal{C} = \Big((-\infty, x_1] \times \ldots \times (-\infty, x_n]\Big)_{\boldsymbol{x} \in \mathbb{R}^n}$ of n-dimensional intervals is a monotone class that generate the Borel σ-field $\mathcal{B}(\mathbb{R}^n)$, so if $F_{\boldsymbol{X}} = F_{\boldsymbol{Y}}$, then $\mu_{\boldsymbol{X}}$ and $\mu_{\boldsymbol{Y}}$ are equal on $\mathcal{C}$. By the monotone class theorem [Kal02, Theorem 1.1], they are equal on $\mathcal{B}(\mathbb{R}^n)$. □

The study of a n-dimensional joint distribution F can be done without reference to a random vector having this joint distribution, but some properties of a joint distribution are more easily expressed in terms of components of a random vector $\boldsymbol{X}$. If not explicitely mentioned, we consider a random vector defined over the canonical probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_F)$.

In order to ease the manipulation of distribution functions, it is useful to introduce the set $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and to extend a given n-dimensional distribution function F to $\overline{\mathbb{R}}^n$ the following way:

Proposition 1.4. *Let F be the distribution function of a real valued n-dimensional random vector, and let $\boldsymbol{x}$ be a vector in $\overline{\mathbb{R}}^n \backslash \mathbb{R}^n$. We denote by:*
- *$\mathcal{I}^-$ the set of indices i such that $x_i = -\infty$,*
- *$\mathcal{I}^+$ the set of indices i such that $x_i = +\infty$,*
- *$\mathcal{I}^0$ the set of indices i such that $x_i \in \mathbb{R}$.*

Then we have $\mathcal{I}^- \cup \mathcal{I}^+ \neq \emptyset$.
- *If $\mathcal{I}^- \neq \emptyset$, then $F(\boldsymbol{x}) = 0$;*
- *If $\mathcal{I}^- = \emptyset$ (and $\mathcal{I}^+ \neq \emptyset$), then $F(\boldsymbol{x}) = \lim_{\forall i \in \mathcal{I}^+, x'_i \to +\infty} F(\boldsymbol{x}')$ with $\forall i \in \mathcal{I}^0, x'_i = x_i$.*

The limit is well-defined thanks to the increasing property of F and the fact that F is bounded by 1.

In the case of a uni-dimensional random vector, also called a **random variable**, the distribution function possesses a generalized inverse called the **quantile function**:

Definition 1.5. Let X be a random variable and F its distribution function. Its **quantile function** F^{-1} is defined by:

$$\forall q \in [0,1],\ F^{-1}(q) = \inf\{x \in \overline{\mathbb{R}} \,|\, F(x) \geq q\}$$

When the distribution function F is increasing, it is invertible and the quantile function is equal to its inverse. The main properties of the quantile function are recalled in the following proposition:

Proposition 1.6. *Let F be a uni-dimensional distribution function and let F^{-1} be the associated quantile function. We have:*

1. *F^{-1} is a left-continuous increasing function from $[0,1]$ into $\overline{\mathbb{R}}$, with the equivalence:*

$$\forall x \in \mathbb{R},\ q \in (0,1],\ F(x) \geq q \Longleftrightarrow x \geq F^{-1}(q)$$

2. *For all $q \in (0,1]$, $F(F^{-1}(q)) \geq q$ with equality if F is continuous at $F^{-1}(q)$;*
3. *Let U be a random variable uniformly distributed on $[0,1]$. Then, F is the distribution function of $F^{-1}(U)$;*
4. *Let X be a random variable with distribution function F. If F is continuous, then $F(X)$ is uniformly distributed over $[0,1]$.*

When studying multivariate probabilistic models, a usual task is to extract multivariate marginal or conditional distribution functions from a given joint distribution function. We start by the definition of marginal random vectors of a given random vector:

Definition 1.7. Let $\boldsymbol{X}$ be a n-dimensional random vector, $\mathcal{I}$ a k-subset of distinct indices $(i_1, \ldots, i_k)$ with $1 \leq i_1 < \ldots < i_k \leq n$ and $\bar{\mathcal{I}} = \{1, \ldots, n\} \backslash \mathcal{I}$. The k-dimensional marginal random vector $\boldsymbol{X}_{\mathcal{I}}$ of $\boldsymbol{X}$ is the random vector defined on the same probability space as $\boldsymbol{X}$ by:

$$\boldsymbol{X}_{\mathcal{I}} = (X_{i_1}, \ldots, X_{i_k})$$

The joint distribution function $F_{\mathcal{I}}$ of $\boldsymbol{X}_{\mathcal{I}}$ is given by:

$$F_{\mathcal{I}} : \left| \begin{array}{ccc} \mathbb{R}^k & \to & [0,1] \\ \boldsymbol{x} & \mapsto & F_{\mathcal{I}}(\boldsymbol{x}) = F_{\boldsymbol{X}}(\tilde{\boldsymbol{x}}) \end{array} \right.$$

where $\tilde{\boldsymbol{x}} \in \overline{\mathbb{R}}^n$ is such that $\forall i_\ell \in \mathcal{I},\ x_{i_\ell} = \tilde{x}_\ell$ and $\forall i_\ell \in \bar{\mathcal{I}},\ x_{i_\ell} = +\infty$.

Two particular cases are of practical importance:
- When $\mathcal{I}$ is reduced to a singleton $\{k\}$, in which case $\boldsymbol{X}_{\mathcal{I}}$ reduces to the random variable X_k and the corresponding marginal joint distribution function is often called the k-th marginal distribution function of $F_{\boldsymbol{X}}$ and is denoted F_k;
- When $\mathcal{I} = \{1, \ldots, k\}$, in which case the marginal distribution function is called the cumulative k-th marginal of $F_{\boldsymbol{X}}$ and is denoted $F_{1,k}$.

The notion of conditional random vector is somewhat more complex. We restrict the definition to continuous random vectors built by conditioning a part of a given continuous random vector by another part of the same random vector:

Definition 1.8. Let $\boldsymbol{X}$ be a n-dimensional absolutely continuous random vector, $\mathcal{I}$ a k-subset of distinct indices $(i_1, \ldots, i_k)$ with $1 \leq i_1 < \ldots < i_k \leq n$ and $\mathcal{J}$ a ℓ subset of distinct indices $(j_1, \ldots, j_\ell)$ with $1 \leq j_1 < \ldots < j_\ell \leq n$ and $\mathcal{I} \cap \mathcal{J} = \emptyset$. The k-conditional

random vector $\boldsymbol{X}_{\mathcal{I}|\mathcal{J}}$ of $\boldsymbol{X}$ given $\boldsymbol{X}_{\mathcal{J}}$ is the random vector defined on the same probability space as $\boldsymbol{X}$, taking values into $\mathbb{R}^k$ and with the following distribution function $F_{\mathcal{I}|\mathcal{J}}$:

$$F_{\mathcal{I}|\mathcal{J}} : \left| \begin{array}{ccc} \mathbb{R}^k & \to & [0,1] \\ \boldsymbol{x} & \mapsto & F_{\mathcal{I}|\mathcal{J}}(\boldsymbol{x}|\tilde{\boldsymbol{x}}) = \left\{ \begin{array}{ll} \dfrac{\partial^\ell F_{\mathcal{I}\cup\mathcal{J}}(\boldsymbol{y})}{\partial\tilde{x}_1,\ldots,\partial\tilde{x}_\ell}/p_{\mathcal{J}}(\tilde{\boldsymbol{x}}) & \text{if } p_{\mathcal{J}}(\tilde{\boldsymbol{x}}) > 0 \\ F_{\mathcal{J}}(\boldsymbol{x}) & \text{otherwise} \end{array} \right. \end{array} \right.$$

where $\tilde{\boldsymbol{x}}$ is a parameter given in $\mathbb{R}^\ell$ and $\boldsymbol{y}$ is such that $y_i = x_s$ if the i-th element of $\mathcal{I}\cup\mathcal{J}$ is the s-th element of $\mathcal{I}$ and $y_i = \tilde{x}_t$ if the i-th element of $\mathcal{I}\cup\mathcal{J}$ is the t-th element of $\mathcal{J}$.

The associated density function $p_{\mathcal{I}|\mathcal{J}}$ is given by:

$$p_{\mathcal{I}|\mathcal{J}} : \left| \begin{array}{ccc} \mathbb{R}^k & \to & \mathbb{R} \\ \boldsymbol{x} & \mapsto & p_{\mathcal{I}|\mathcal{J}}(\boldsymbol{x}|\tilde{\boldsymbol{x}}) = \left\{ \begin{array}{ll} p_{\mathcal{I}\cup\mathcal{J}}(\boldsymbol{y})/p_{\mathcal{J}}(\tilde{\boldsymbol{x}}) & \text{if } p_{\mathcal{J}}(\tilde{\boldsymbol{x}}) > 0 \\ p_{\mathcal{J}}(\boldsymbol{x}) & \text{otherwise} \end{array} \right. \end{array} \right.$$

A special case of conditional random vector will play a major role in the definition of the Rosenblatt transformation. It corresponds to $\mathcal{I} = \{k\}$ and $\mathcal{J} = \{1,\ldots,k-1\}$, with $k \in \{2,\ldots,n\}$. In this case, the conditional random vector reduces to the random variable $X_k|X_1 = x_1,\ldots,X_{k-1} = x_{k-1}$, with a distribution function $F_{k|1,\ldots,k-1}$ given by:

$$F_{k|1,\ldots,k-1} : \left| \begin{array}{ccc} \mathbb{R} & \to & [0,1] \\ x_k & \mapsto & F_{k|1,\ldots,k-1}(x_k|x_1,\ldots,x_{k-1}) = \\ & & \left\{ \begin{array}{ll} \dfrac{\partial^{k-1}F_{1,k}(x_1,\ldots,x_k)}{\partial x_1\ldots\partial x_{k-1}}/\dfrac{\partial^{k-1}F_{1,k-1}(x_1,\ldots,x_{k-1})}{\partial x_1\ldots\partial x_{k-1}} & \text{if } \dfrac{\partial^{k-1}F_{1,k-1}(x_1,\ldots,x_{k-1})}{\partial x_1\ldots\partial x_{k-1}} > 0 \\ F_{1,k}(\boldsymbol{x}) & \text{otherwise} \end{array} \right. \end{array} \right. \tag{1.1}$$

1.1 Copulas

The main source for this section is [Nel06]. It is focused on the notion of **copula**, that plays a central role in the modeling of stochastic dependence.

We start by the definition of a copula, as found in [Nel06, Definitions 2.10.5 and 2.10.6]:

Definition 1.9. A copula is a function C defined on $[0,1]^n$, taking value into $[0,1]$ and verifying:

1. For all $\boldsymbol{u} \in [0,1]^n$ with at least one component equal to 0, $C(\boldsymbol{u}) = 0$ (C is **grounded**);
2. C is n-increasing: $\forall \boldsymbol{a}, \boldsymbol{b} \in [0,1]^n$,

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} C(u_1^{i_1},\ldots,u_n^{i_n}) \geq 0 \tag{1.2}$$

   with $\forall j \in \{1,\ldots,n\}$, $u_j^1 = a_j$ and $u_j^2 = b_j$.
3. For all $\boldsymbol{u} \in [0,1]^n$ with $\forall i \in \{1,\ldots,n\}\backslash\{k\}, u_i = 1$:

$$C(\boldsymbol{u}) = u_k \tag{1.3}$$

A first link between copulas and joint distribution functions is given in the following theorem:

Theorem 1.10. *There is a one-to-one mapping between the n-dimensional copulas and the restriction to $[0,1]^n$ of the distribution functions of the n-dimensional random vectors with one-dimensional marginal distributions uniform over $[0,1]$.*

Proof. If F is the distribution function of a n-dimensional random vector $\boldsymbol{U}$ with one-dimensional marginal distributions uniform over $[0,1]$, then its restriction $\tilde{F}$ to $[0,1]^n$ is such that:

1. For all $\boldsymbol{u} \in [0,1]^n$ with at least one component x_i equal to 0,

$$\tilde{F}(\boldsymbol{u}) = \mathbb{P}\left(U_1 \leq u_1, \ldots, U_n \leq u_n\right) \leq \mathbb{P}\left(U_i \leq u_i\right) = 0$$

   so $\tilde{F}$ is grounded.

2. If $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^n$ are such that $\boldsymbol{a} \leq \boldsymbol{b}$, we note by $\mathbb{1}_{(\boldsymbol{a},\boldsymbol{b}]} = \prod_{j=1}^{n} \mathbb{1}_{(a_j,b_j]}$ the characteristic function of the interval $(\boldsymbol{a},\boldsymbol{b}]$. We have:

$$\mathbb{1}_{(\boldsymbol{a},\boldsymbol{b}]} = \sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} \prod_{j=1}^{n} 1_{(0,u_j^{i_j}]}$$

   with $\forall j \in \{1,\ldots,n\}$, $u_j^1 = a_j$ and $u_j^2 = b_j$. Taking the expectation with respect to $\boldsymbol{U}$, as $\{U_k \leq u_k\} = \{0 < U_k \leq u_k\}$ up to a negligible set, we get:

$$\mathbb{P}\left(U \in (\boldsymbol{a},\boldsymbol{b}]\right) = \sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} \tilde{F}(u_1^{i_1}, \ldots, u_n^{i_n}) \geq 0$$

   with $\forall j \in \{1,\ldots,n\}$, $u_j^1 = a_j$ and $u_j^2 = b_j$.

3. As F has uniform one-dimensional marginal distributions over $[0,1]$, then for all $\boldsymbol{x} \in [0,1]^n$ with $\forall i \in \{1,\ldots,n\}\backslash\{k\}, x_i = 1$, $F(\boldsymbol{x}) = F_k(x_k) = x_k$.

so $\tilde{F}$ is a copula.

Conversely, if C is a copula, to show that C is also the restriction to $\Omega = [0,1]^n$ of a joint distribution function, we proceed as follows:

- We define a function P defined on the intervals of the form $(\boldsymbol{0},\boldsymbol{u}]$ for all $\boldsymbol{u} \in [0,1]^n$ by $P((\boldsymbol{0},\boldsymbol{u}]) = C(\boldsymbol{u})$, with $P(\emptyset) = 0 = C(\boldsymbol{u})$ if $\boldsymbol{u}$ has a component equal to 0.
- We extend P to the family $\mathcal{A}$ of finite unions of intervals $(\boldsymbol{a},\boldsymbol{b}]$ using the n-increasing property of C.
- The family $\mathcal{A}$ generates the Borelian σ-field of $(0,1]^n$, and as $P((0,1]^n) = C(\boldsymbol{1}) = 1$, P can be extended to a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{B}(\Omega))$ to get the probability space $\mathcal{E} = (\Omega, \mathcal{B}(\Omega), \mathbb{P})$. If we consider the identity function on $\mathcal{E}$, it defines a random vector $\boldsymbol{U}$ and by construction, its joint distribution function is C.

□

This theorem allows to see a given copula as a joint distribution function as far as it is not evaluated on vectors outside of $[0,1]^n$.

The next theorem, due to Sklar [Skl59], is a central result in the theory of copulas as it fully explains the link between copulas and joint distribution functions.

Theorem 1.11 (Sklar, 1959)**.** *Let F be a n-dimensional distribution function whose marginal distribution functions are $F_1, \ldots, F_n$. There exists a copula C of dimension n such that for $\boldsymbol{x} \in \overline{\mathbb{R}}^n$, we have:*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)). \tag{1.4}$$

If the marginal distributions $F_1, \ldots, F_n$ are continuous, the copula C is unique; otherwise, it is uniquely determined on $F_1(\overline{\mathbb{R}}) \times \ldots \times F_n(\overline{\mathbb{R}})$. Conversely, if C is a copula of dimension n and $F_1, \ldots, F_n$ are n univariate distribution functions, then the function F defined by (1.4) is a multivariate distribution function of dimension n with marginal distribution functions $F_1, \ldots, F_n$.

In the case of continuous marginal distributions, for all $\boldsymbol{u} \in [0,1]^n$, we have:

$$C(\boldsymbol{u}) = F({F_1}^{(-1)}(u_1), \ldots, {F_n}^{(-1)}(u_n)) \tag{1.5}$$

and if F is absolutely continuous with density p, so is C and the two density functions are linked by:

$$p(\boldsymbol{x}) = c(F_1(x_1), \ldots, F_n(x_n)) \prod_{i=1}^{n} p_i(x_i) \tag{1.6}$$

where c is the density function of C and p_i is the density function of X_i.

In the continuous marginal case, one can see that there is a one-to-one correspondence between the joint distribution and the set of marginal distributions and copula. In this sense, the copula is exactly what remains of the dependence structure of the joint distribution once the effect of the marginal distributions has been filtered out.

It can be interpreted in several ways in a modeling perspective. The first way is the synthesis way: one want to build a joint distribution function given marginal distribution functions and potentially some partial information about the interactions between the marginals. It is a matter of copula selection, potentially under constraints. The second way is the analysis way: given a joint distribution function or a sampling procedure, one want to separate the effects of the marginal distributions and the dependence structure in the probabilistic behaviour of such a multivariate model.

There is a link between marginal distribution functions and marginal copulas, as presented in the following proposition.

Proposition 1.12. *Let $\boldsymbol{X}$ be a continuous random vector with a joint distribution function F defined by its copula C and its marginal distribution functions $F_1, \ldots, F_n$. The distribution function $F_{1,k}$ of the cumulated k-th marginal random vector $\boldsymbol{X}_{1,k}$ is linked to C defined by its marginal distributions F_i and the copula $C_{1,k}$ through the relation:*

$$F_{1,k}(x_1, \ldots, x_k) = C_{1,k}(F_1(x_1), \ldots, F_k(x_k)) \tag{1.7}$$

with

$$C_{1,k}(u_1, \ldots, u_k) = C(u_1, \ldots, u_k, 1, \ldots, 1) \tag{1.8}$$

Proposition 1.13. *Let $\boldsymbol{X}$ be an absolutely continuous random vector with a distribution defined by its copula C and its marginal distribution functions F_i. The distribution function*

of the conditional variable $X_k|X_1,\ldots,X_{k-1}$ is defined by its marginal distributions F_i and the copula $C_{k|1,\ldots,k-1}$ through the relation:

$$F_{k|1,\ldots,k-1}(x_k|x_1,\ldots,x_{k-1}) = C_{k|1,\ldots,k-1}(F_k(x_k)|F_1(x_1),\ldots,F_{k-1}(x_{k-1})) \tag{1.9}$$

with

$$C_{k|1,\ldots,k-1}(u_k|u_1,\ldots,u_{k-1}) = \frac{\partial^{k-1}C_{1,k}(u_1,\ldots,u_k)}{\partial u_1\ldots\partial u_{k-1}} \Big/ \frac{\partial^{k-1}C_{1,k-1}(u_1,\ldots,u_{k-1})}{\partial u_1\ldots\partial u_{k-1}} \tag{1.10}$$

As a matter of fact, relation (1.10) is the direct application of Proposition 1.1 to the distribution C. Furthermore, Definition 1.1 and relation (1.4) lead to:

$$\begin{aligned} F_{k|1,\ldots,k-1}(x_k|x_1,\ldots,x_{k-1}) &= \left[\prod_{i=1}^{k-1} p_i(x_i)\right] \frac{\partial^{k-1}C_{1,k}(F_1(x_1),\ldots,F_k(x_k))}{\partial u_1\ldots\partial u_{k-1}} \Big/ \ldots \\ &\ldots \left[\prod_{i=1}^{k-1} p_i(x_i)\right] \frac{\partial^{k-1}C_{1,k-1}(F_1(x_1),\ldots,F_k(x_{k-1}))}{\partial u_1\ldots\partial u_{k-1}} \end{aligned} \tag{1.11}$$

where p_i is the probability density function of X_i.

A copula is Lipschitz on $[0,1]^n$ by [Nel06, Theorem 2.10.7]:

Theorem 1.14. *Let C be a n-dimensional copula. Then for every $\boldsymbol{u}$ and $\boldsymbol{v}$ in $[0,1]^n$,*

$$|C(\boldsymbol{v}) - C(\boldsymbol{u})| \leq \sum_{k=1}^{n} |v_k - u_k| \tag{1.12}$$

In particular, the continuity of the marginal distributions of a copula implies that there is no atom (ie mass point) in the associated distribution.

We give in Table 1.1 some classical bi-dimensional copulas that will be used in the sequel of the manuscript for numerical applications.

Several key properties of copulas will be used in the sequel of the manuscript. The first one is the Fréchet-Hoeffding bounds [Nel06, Theorem 2.10.12]:

Theorem 1.15. *Let C be a n-dimensional copula. Then $\forall \boldsymbol{u} \in [0,1]^n$, we have:*

$$W_n(\boldsymbol{u}) = \max(u_1 + \ldots + u_n - 1, 0) \leq C(\boldsymbol{u}) \leq M_n(\boldsymbol{u}) = \min(u_1,\ldots,u_n) \tag{1.13}$$

These bounds are tight, in the sense that for any point $\boldsymbol{u} \in [0,1]^n$ and any side of the inequality, there exists a copula C such that this inequality is indeed an equality. The lower bound W_n is a copula if and only if $n = 2$, while the upper bound M_n is always a copula called the **min** *copula.*

The min copula has the following probabilistic interpretation, [Nel06, Theorem 2.10.14]:

Theorem 1.16. *For $n \geq 2$, each of the components $X_1,\ldots,X_n$ of a continuous random vector $\boldsymbol{X}$ is almost surely a strictly increasing function of any of the others if and only if the copula of $\boldsymbol{X}$ is M_n. In this case, there exist $n-1$ almost surely strictly increasing functions $\phi_2,\ldots,\phi_n$ such that $\forall i \in \{2,\ldots,n\}$, $X_i = \phi_i(X_1)$.*

Name	$C(u_1, u_2)$	Parameter
Independent	$u_1 u_2$	
Normal	$\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) \mathrm{d}s\,\mathrm{d}t$	$\lvert\rho\rvert < 1$
Student	$\int_{-\infty}^{T_\nu^{-1}(u_1)} \int_{-\infty}^{T_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2} \mathrm{d}s\,\mathrm{d}t$	$\lvert\rho\rvert < 1$, $\nu > 0$
Frank	$-\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1}{e^{-\theta} - 1}\right)$	$\theta \in \mathbb{R}^*$
Clayton	$\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$	$\theta \in [-1, +\infty) \backslash \{0\}$
Comp. Clayton	$u_1 + u_2 - 1 + C_{Clayton}(1 - u_1, 1 - u_2)$	
Gumbel	$\exp\left(-\left((-\log(u_1))^\theta + (-\log(u_2))^\theta\right)^{1/\theta}\right)$	$\theta \in [1, +\infty)$

Table 1.1: Examples of usual bi-dimensional copulas. Φ^{-1} (resp. T_ν^{-1}) stands for the quantile function of the standard normal (resp. Student-ν) distribution.

In the bi-dimensional case, the function W_2 is also a copula, which has the following interpretation, [Nel06, Theorem 2.5.5]:

Theorem 1.17. *For a bi-dimensional continuous random vector $\boldsymbol{X} = (X_1, X_2)$, each of its components is almost surely a strictly decreasing function of the other one if and only if the copula of $\boldsymbol{X}$ is W_2.*

The next proposition plays a central role in the study of iso-probabilistic transforms in link with copulas:

Proposition 1.18.
If $\boldsymbol{X}$ has as a joint distribution with copula C and if $(\alpha_1, \ldots, \alpha_n)$ are n almost everywhere strictly increasing functions defined respectively on the supports of the X_i, then C is also the copula of $(\alpha_1(X_1), \ldots, \alpha_n(X_n))$.

1.2 Dependence through scalars

Several attempts to quantify the stochastic dependence between two random variables through a scalar value have been made. General classes of scalar measures have been defined so far, namely the **measures of concordance**, the **measures of dependence** and the **measures of association**, in order to describe the intensity of the stochastic relationship that links two random variables. Given these notions, it is interesting to review the most widely used quantities in order to check whether they are able to properly summarize stochastic dependence.

A more detailed presentation of these quantities as well as many other ones can be found in [Nel06], [Joe97] or [ELM03].

1.2.1 Measures of concordance, dependence and association

The first notion that summarizes some dependence information between two random variables with a scalar is the notion of **measure of concordance**, introduced for the first time in [Sca84] and recalled in [Nel06].

Definition 1.19. A **measure of concordance** κ between the two components X_1 and X_2 of a bi-dimensional random vector $\boldsymbol{X}$ is a scalar function of its distribution function that has the following properties:

1. κ is defined for all continuous bi-dimensional random vectors $\boldsymbol{X}$,
2. $\kappa(X_1, X_2) = \kappa(X_2, X_1)$,
3. κ depends only on the copula $C_{\boldsymbol{X}}$ of $\boldsymbol{X}$ and is monotone in $C_{\boldsymbol{X}}$, it means that if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two bi-dimensional random vectors with respective copulas $C_{\boldsymbol{X}}$ and $C_{\boldsymbol{Y}}$ and if $\forall \boldsymbol{u} \in [0,1]^2, C_{\boldsymbol{X}}(\boldsymbol{u}) \geq C_{\boldsymbol{Y}}(\boldsymbol{u})$, then $\kappa(X_1, X_2) \geq \kappa(Y_1, Y_2)$.
4. $\kappa(X_1, X_2) \in [-1, 1]$, $\kappa(X_1, X_1) = 1$,
5. If X_1 and X_2 are independent, then $\kappa(X_1, X_2) = 0$,
6. $\kappa(X_1, -X_2) = \kappa(-X_1, X_2) = -\kappa(X_1, X_2)$,
7. If C_n is a sequence of copulas that converges pointwise to the copula C, then $\kappa(C_n)$ converges pointwise to $\kappa(C)$, where $\kappa(C)$ is a shorthand for $\kappa(X_1, X_2)$, where the support of $\boldsymbol{X}$ is $[0,1]^2$ and its distribution function restricted to this support is C.

Examples of such measures are the Spearman rho or the Kendall tau [Nel06, Theorem 5.1.8], defined in the next section. There is a visible link with the linear correlation coefficient, as the next theorem shows [Sca84, Theorem 3]:

Theorem 1.20. *Let $\boldsymbol{X}$ be a bi-dimensional Gaussian vector with correlation coefficient r, then for any measure of concordance κ, κ is an increasing function of r.*

The second notion is the notion of **measure of dependence**, where the aim is also to summarize some dependence information between two random variables but with a normalization that allows to interpret it as a probability, with less symmetry constraints but with a stronger meaning for the zero value.

Definition 1.21. A **measure of dependence** δ between the two components X_1 and X_2 of a bi-dimensional random vector $\boldsymbol{X}$ is a scalar function of its distribution function that has the following properties:

1. δ is defined for all continuous bi-dimensional random vectors $\boldsymbol{X}$,
2. $\delta(X_1, X_2) = \delta(X_2, X_1)$,
3. $\delta(X_1, X_2) \in [0, 1]$
4. X_1 and X_2 are independent if and only if $\delta(X_1, X_2) = 0$,
5. $\delta(X_1, X_2) = 1$ if and only if $X_2 = \phi(X_1)$ a.s., where ϕ is strictly monotone,
6. if g and h are almost everywhere strictly increasing functions defined respectively over the support of X_1 and the support of X_2, then $\delta(g(X_1), h(X_2)) = \delta(X_1, X_2)$,
7. If C_n is a sequence of copulas that converge pointwise to the copula C, then $\delta(C_n)$ converges pointwise to $\delta(C)$, where $\delta(C)$ is a shortcut for $\delta(X_1, X_2)$, where the support of $\boldsymbol{X}$ is $[0, 1]^2$ and its distribution function restricted to this support is C.

Remark 1.22. The fourth property is very strong: it allows to characterize the independence between two random variables through a scalar value.

Remark 1.23. The sixth property implies that a measure of dependence is a function of the copula of $\boldsymbol{X}$ and not of its marginal distribution functions: as F_1 and F_2 are almost everywhere strictly increasing over the support of X_1 and the support of X_2, $\delta(W_1, W_2) = \delta(X_1, X_2)$ where $\boldsymbol{W}$ is the random vector defined by $\boldsymbol{W} = (F_1(X_1), F_2(X_2))$. By Proposition 1.18, $\boldsymbol{X}$ and $\boldsymbol{W}$ share the same copula C, and the distribution function of $\boldsymbol{W}$ is exactly C, so $\delta(X_1, X_2)$ is a function of C only.

Remark 1.24. A measure of dependence is never a measure of concordance, due to the symmetry constraint in the definition of measures of concordance (points 4 and 6), which is not compatible with the nonnegativity of measures of dependence.

The following exemples of measures of dependence are taken from [SW81].

Example 1.25. Let $\boldsymbol{X} = (X_1, X_2)$ be a continuous bi-dimensional random vector with copula C. The following measures are measures of dependence:

$$\sigma(X_1, X_2) = 12 \iint_{[0,1]^2} |C(u, v) - uv| \, \mathrm{d}u\mathrm{d}v$$

$$\gamma(X_1, X_2) = \left(90 \iint_{[0,1]^2} (C(u, v) - uv)^2 \, \mathrm{d}u\mathrm{d}v \right)^{1/2}$$

These measures are respectively derived from the 1-distance and the 2-distance between a given bidimensional copula and the bidimensional independent copula. The constants are such that the range of each measure is exactly $[-1, 1]$ over the set of bidimensional copulas, obtained when $C = M_2$ is the bidimensional min copula defined in Theorem 1.15.

Such a construction can be extended to any L^p distance for $1 \leq p < +\infty$, the measure obtained for $p = +\infty$ satisfying all the properties of a measure of dependence excepted the property (5).

The third notion is the weakest notion to summarize dependence information into a scalar value.

Definition 1.26. A **measure of association** r between the two components X_1 and X_2 of a bi-dimensional random vector $\boldsymbol{X}$ is a scalar function of its distribution function that has the following properties:

1. $-1 \leq r(X_1, X_2) \leq 1$
2. If X_1 and X_2 are independent, then $r(X_1, X_2) = 0$
3. If g and h are almost everywhere strictly increasing functions defined respectively over the support of X_1 and the support of X_2, then $r(g(X_1), h(X_2)) = r(X_1, X_2)$,

The aim of this definition is to cover any reasonable scalar function linked to the dependence between the components of a random vector, with two normalization constraints that allow to detect possible independence and to compare the dependence intensity on a fixed scale.

Remark 1.27. The third property shows that a measure of association is a function of the copula of $\boldsymbol{X}$ only, see the remark related to the sixth property of a measure of dependence.

Remark 1.28. Measures of concordance and measures of dependence are both measures of association, but a measure of association can be neither a measure of concordance nor a measure of dependence. An example of such a measure is the following:

1. Take a measure of concordance κ
2. For all continuous bi-dimensional random vector $\boldsymbol{X}$, define the measure of association $\tilde{\kappa}$ by:

$$\tilde{\kappa}(X_1, X_2) = \begin{cases} \kappa(X_1, X_2) & \text{if } \kappa(X_1, X_2) \geq 0 \\ \dfrac{1}{2}\kappa(X_1, X_2) & \text{otherwise} \end{cases}$$

It satisfies the definition of a measure of association, but is neither a measure of dependence nor a measure of concordance as $\hat{\kappa}(X_1, -X_1) = -1/2$ is negative but not equal to -1.

1.2.2 Overview of classical candidates for measures of association

In this section, we review four candidates as measure of association, namely the linear correlation, the Spearman rho, the Kendall tau and the (upper and lower) coefficient of tail dependence, and show that with the exception of the linear correlation, the three other candidates are proper measures of association. In the sequel, we will restrict ourselves to **continuous** random vectors.

The use of **linear correlation** coefficient to describe dependence is largely spread in industrial studies. It is mainly because it is quite easy to estimate and because it appears

to be a natural measure of dependence for Gaussian vectors, which is the most widely used model in the case of correlated variables.

Definition 1.29. Let $\boldsymbol{X} = (X_1, X_2)$ be a random vector with finite and positive variance, with marginal distribution functions F_1 and F_2 and copula C. The **linear correlation** $\rho(\boldsymbol{X})$ between X_1 and X_2 is given by:

$$\rho(\boldsymbol{X}) = \mathbb{E}\left[\left(\frac{X_1 - \mathbb{E}[X_1]}{\sqrt{\mathbf{Var}[X_1]}}\right)\left(\frac{X_2 - \mathbb{E}[X_2]}{\sqrt{\mathbf{Var}[X_2]}}\right)\right]$$

In terms of the marginal distributions and copula, it is given by:

$$\rho(\boldsymbol{X}) = \frac{1}{\sqrt{\mathbf{Var}[X_1]}\sqrt{\mathbf{Var}[X_2]}} \iint_{[0,1]^2} [C(u,v) - uv]\, \mathrm{d}F_1^{-1}(u)\mathrm{d}F_2^{-1}(v) \tag{1.14}$$

The linear correlation coefficient does **not** fullfil point (3) of the definition of a measure of association. For example, if we consider a bidimensional Gaussian vector $\boldsymbol{X} = (X_1, X_2)$ with standard marginal distributions and linear correlation $\rho(\boldsymbol{X}) \neq 0$, then the random vector $\boldsymbol{Y} = (X_1, X_2|X_2|)$ should have the same linear correlation as $\boldsymbol{X}$ if the linear correlation coefficient would be a measure of association, but the linear correlation between X_1 and $X_2|X_2|$ is equal to:

$$\rho(\boldsymbol{Y}) = \frac{2\sqrt{2}}{\pi\sqrt{3}}\rho(\boldsymbol{X})\sqrt{1 - \rho(\boldsymbol{X})^2}$$

which is different from $\rho(\boldsymbol{X})$.

Relation (1.14) shows that the linear correlation depends both on the copula and the marginal distributions. Thus, it is not possible to set its value independently of the determination of the marginal distributions.

The following Fréchet-Hoeffding theorem [KC06, Theorem 3.1] discusses the link between the linear correlation coefficient and the marginals.

Theorem 1.30. *Let $\boldsymbol{X} = (X_1, X_2)$ be a bi-dimensional random vector with given marginal distribution functions F_1 and F_2. The possible values for the linear correlation coefficient $\rho(\boldsymbol{X})$ form an interval $[\rho_{min}, \rho_{max}]$ included in $[-1, 1]$ that depends on F_1 and F_2, the inclusion being strict in the general case. The lower bound corresponds to the W_2 copula and the upper bound to the M_2 copula.*

An illustration of this fact is given by the following example, taken from [ELM03]. Let X_1 be log-normaly distributed with $\mathbb{E}[\log(X_1)] = 0$ and $\mathbf{Var}[\log(X_1)] = 1$, and X_2 be log-normaly distributed with $\mathbb{E}[\log(X_2)] = 0$ and $\mathbf{Var}[\log(X_2)] = \sigma^2$. Then, $\rho(\boldsymbol{X}) \in [\rho_{min}, \rho_{max}] \subsetneq [-1, 1]$ with:

$$\begin{cases} \rho_{min} &= \dfrac{e^{-\sigma} - 1}{\sqrt{e-1}\sqrt{e^{\sigma^2} - 1}} \\ \rho_{max} &= \dfrac{e^{\sigma} - 1}{\sqrt{e-1}\sqrt{e^{\sigma^2} - 1}} \end{cases} \tag{1.15}$$

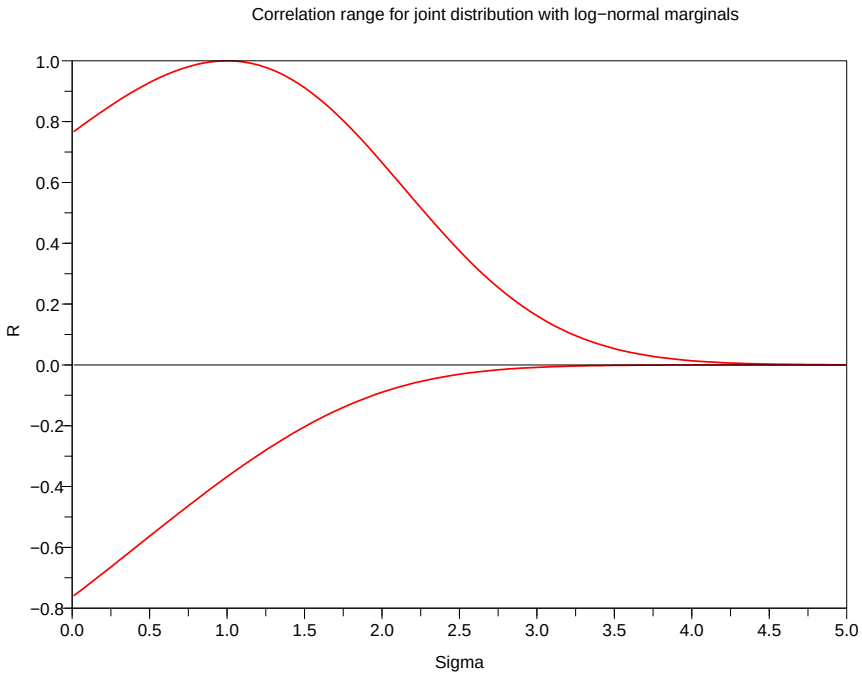


Figure 1.1: Fréchet bounds for $R = \rho(\boldsymbol{X})$ as a function of σ.

On Figure 1.1, we see that $\lim_{\sigma\to+\infty} \rho_{min} = \lim_{\sigma\to+\infty} \rho_{max} = 0$: even if X_1 and X_2 are linked by a strictly monotoneous relation, their linear correlation coefficient can be arbitrarily close to zero!

The Spearman rho, also named **rank correlation**, and **the Kendall tau** are another well-known quantities used to quantify the dependence between X_1 and X_2. The Spearman rho is defined as the linear correlation between the **ranks** of X_1 and X_2, i.e. when X_1 and X_2 have been transformed through their respective distribution functions:

Definition 1.31. Let $\boldsymbol{X} = (X_1, X_2)$ be a random vector with marginal distribution functions F_1 and F_2 and copula C. Its **rank correlation** $\rho_S(\boldsymbol{X})$, also named its **Spearman rho**, is defined by:

$$\begin{aligned}\rho_S(\boldsymbol{X}) &= \rho(F_1(X_1), F_2(X_2)) \\ &= 12 \iint_{[0,1]^2} uv\, dC(u,v) - 3 \\ &= 12 \iint_{[0,1]^2} C(u,v)\, dudv - 3 \end{aligned} \tag{1.16}$$

while the Kendall tau measures the concordance between the two components of a bi-dimensional random vector $\boldsymbol{X} = (X_1, X_2)$:

Definition 1.32. Let $\boldsymbol{X} = (X_1, X_2)$ be a random vector with marginal distribution functions F_1 and F_2 and copula C, and (X_1', X_2') be an independent copy of (X_1, X_2). The Kendall tau $\tau(\boldsymbol{X})$ is the difference between the probability of concordance and the probability of discordance between X_1 and X_2:

$$\tau(\boldsymbol{X}) = \mathbb{P}\left((X_1 - X_1')(X_2 - X_2') > 0\right) - \mathbb{P}\left((X_1 - X_1')(X_2 - X_2') < 0\right)$$

The Kendall tau can be also computed using one of the following equivalente relations:

$$\begin{aligned}\tau(\boldsymbol{X}) &= \mathbb{E}\left[\operatorname{sgn}(X_1 - X_1')\operatorname{sgn}(X_2 - X_2')\right] \\ &= 4\,\mathbb{E}\left[C(F_1(X_1), F_2(X_2))\right] - 1 \\ &= 4\iint_{[0,1]^2} C(u,v)\,dC(u,v) - 1\end{aligned} \tag{1.17}$$

where the sign function sgn is defined by:

$$\forall x \in \mathbb{R},\ \operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

The Spearman rho and the Kendall tau are measures of concordance. They are easy to estimate in a robust way (i.e not sensitive to outliers), and can be linked to the linear correlation ρ by the following formulas in the case of Gaussian vectors :

$$\begin{cases} \rho_S &= \dfrac{6}{\pi}\arcsin\left(\dfrac{1}{2}\rho\right) \\ \tau &= \dfrac{2}{\pi}\arcsin\left(\rho\right) \end{cases} \tag{1.18}$$

Equations (1.16) and (1.17) allow to compute all the Spearman rho or the Kendall tau between the components of a random vector, but in practice the problem is more to find a copula compatible with a given set of Spearman rho or Kendall tau that have been estimated from data. The next theorem shows that it is always possible to find such a copula for small dimensions (see [KC06, Theorem 4.4]):

Theorem 1.33. *Let $\boldsymbol{M}$ be a symmetric positive definite matrix of dimension n with unit diagonal. If $n = 2$ or $n = 3$, there exists a copula such that its matrix of Spearman rho is $\boldsymbol{M}$. The same result holds if $\boldsymbol{M}$ is seen as a matrix of Kendall tau. No such result is known for $n > 3$.*

This result contrasts with the case of the linear correlation for which the compatibility condition involves both the copula and the marginal distributions. The case of small dimensions is much more useful than it seems at first glance: in real-life applications, even if several dozens of variables are involved, most of the time they are independent by blocks, each block involving only a small number of variables.

The **coefficients of tail dependence** are measures of association that aim at quantifying the dependence between random variables when they take simultaneously extreme values. These coefficients are based on the **tail copulas** of a random vector, and are an important part of the **multi-dimentional theory of extrems** (see [KN00]). These tail copulas are defined by:

Definition 1.34. Let $\boldsymbol{X}$ be a n-dimensional random vector with distribution function F. If for the subsets $\mathcal{I}, \mathcal{J} \subset \{1, \ldots, n\}$, $\mathcal{I} \cap \mathcal{J} = \emptyset$, the following limit exists everywhere on $[0,1]^n$:

$$\Lambda_U^{\mathcal{I},\mathcal{J}}(\boldsymbol{u}) = \lim_{t\to\infty} \mathbb{P}\left(F_i(X_i) > 1 - \frac{u_i}{t},\ \forall i \in \mathcal{I} \,\middle|\, F_j(X_j) > 1 - \frac{u_j}{t},\ \forall j \in \mathcal{J}\right) \tag{1.19}$$

then the function $\Lambda_U^{\mathcal{I},\mathcal{J}} : [0,1]^n \to [0,1]$ is called an upper tail-copula associated with F with respect to $\mathcal{I}, \mathcal{J}$.

The corresponding lower tail-copula is defined by:

$$\Lambda_L^{\mathcal{I},\mathcal{J}}(\boldsymbol{u}) = \lim_{t\to\infty} \mathbb{P}\left(F_i(X_i) \leq \frac{u_i}{t}, \forall i \in \mathcal{I} \,\middle|\, F_j(X_j) \leq \frac{u_j}{t}, \forall j \in \mathcal{J}\right) \tag{1.20}$$

provided the limit exists.

In the bidimensional case, we take $\mathcal{I} = \{1\}$ and $\mathcal{J} = \{2\}$ and we denote by $\Lambda_U(u_1, u_2) = \Lambda_U^{\{1\},\{2\}}(\boldsymbol{u})$ and $\Lambda_L(u_1, u_2) = \Lambda_L^{\{1\},\{2\}}(\boldsymbol{u})$. The coefficients of tail dependence are then defined by:

Definition 1.35. Let $\boldsymbol{X} = (X_1, X_2)$ be a bi-dimensional random vector with marginal distribution functions F_1 and F_2, and copula C. The **coefficients of upper and lower tail dependence** $\lambda_U(\boldsymbol{X})$ and $\lambda_L(\boldsymbol{X})$ are defined as :

$$\begin{cases} \lambda_U(\boldsymbol{X}) &= \lim_{q\to 1^-} \mathbb{P}\left(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q)\right) = \Lambda_U(1,1) \\ \lambda_L(\boldsymbol{X}) &= \lim_{q\to 0^+} \mathbb{P}\left(X_2 < F_2^{-1}(q) | X_1 < F_1^{-1}(q)\right) = \Lambda_L(1,1) \end{cases} \tag{1.21}$$

provided that these limits $\lambda_U \in [0,1]$ and $\lambda_L \in [0,1]$ exist. If F_1 and F_2 are continuous, we also have:

$$\begin{cases} \lambda_U(\boldsymbol{X}) &= \lim\limits_{q\to 1^-} \left(2 - \dfrac{1 - C(q,q)}{1-q}\right) \\ \lambda_L(\boldsymbol{X}) &= \lim\limits_{q\to 0^+} \dfrac{C(q,q)}{q} \end{cases} \tag{1.22}$$

In other words, the coefficient of upper (resp. lower) tail dependence is the probability that the random variable X_2 exceeds (resp. remains under) its quantile of order q, knowing that X_1 exceeds (resp. remains under) its quantile of the same order, when this order tends to 1 (resp. 0) : it is clearly an indicator of the dependence for the upper (resp. lower) extremes.

Relation (1.22) shows that the existence and the value of the coefficient of upper or lower tail dependence are properties of the copula only : they are proper measures of association. We remark that the existence of a coefficient of upper tail dependence does not imply the existence of a coefficient of lower tail dependence.

As this coefficient can play a very important role in reliability studies, we give three examples of 2D copulas and their coefficients of upper and lower tail dependence:

– Normal copula with correlation $|\rho| < 1$:

$$\lambda_L(\boldsymbol{X}) = \lambda_U(\boldsymbol{X}) = 0 \tag{1.23}$$

– Student copula with correlation $|\rho| < 1$ and ν degrees of freedom (see Table 1.1):

$$\lambda_L(\boldsymbol{X}) = \lambda_U(\boldsymbol{X}) = 2\left(1 - T_{\nu+1}\left(\sqrt{(\nu+1)(1-\rho)/(1+\rho)}\right)\right) \tag{1.24}$$

where $T_{\nu+1}$ is the density function of the one-dimensional Student distribution with $\nu+1$ degrees of freedom defined by $\forall x \in \mathbb{R}, T_{\nu+1}(x) = \dfrac{1}{\sqrt{2\pi}}\left(1 + \dfrac{x^2}{\nu+1}\right)^{-(\nu+3)/2}$.

– Clayton copula with parameter $\theta \geq 0$ (see Table 1.1):

$$\lambda_U(\boldsymbol{X}) = 0,\ \lambda_L(\boldsymbol{X}) = 2^{-1/\theta} \tag{1.25}$$

The presence or not of positive tail dependence will certainly play a major role in reliability analysis, as we have in the case of an upper tail dependence $\lambda_U > 0$ [1]:

$$\mathbb{P}\left(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\right) = \lambda_U(1-q) + o(1-q) \text{ as } q \to 1 \tag{1.26}$$

In the absence of upper tail dependence, also called **upper tail independence**, we have:

$$\mathbb{P}\left(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\right) = o(1-q) \text{ as } q \to 1 \tag{1.27}$$

It can be divided into 3 sub-cases:

– The **perfect independence** which corresponds to:

$$\mathbb{P}\left(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\right) = \Theta\left((1-q)^2\right) \text{ as } q \to 1 \tag{1.28}$$

– The **independence with positive association** which corresponds to:

$$\mathbb{P}\left(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\right) = \omega\left((1-q)^2\right) \text{ as } q \to 1 \tag{1.29}$$

– The **independence with negative association** which corresponds to:

$$\mathbb{P}\left(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\right) = o\left((1-q)^2\right) \text{ as } q \to 1 \tag{1.30}$$

If the failure domain has a significant part in the positive quadrant $\mathcal{Q}^+ = \{\boldsymbol{x} \in \mathbb{R}^n \,|\, \forall i \in \{1,\dots,n\},\ x_i \geq 0\}$, one can expect that in the case of upper tail independence, the probability of failure will be negligible compared to the case of positive upper tail dependence.

The difference between the case of upper tail independence and upper tail perfect independence is that in the independent case, **negligible** means of one order of magnitude smaller, whereas upper tail independence says nothing about the effective rate of decrease: it can be much slower (independence with positive association) or much faster (independence with negative association) than the perfect independent case.

Remark 1.36. When X_1 and X_2 are independent, we are in the case of perfect independence as in this case, $\forall q \in (0,1)$, $\mathbb{P}\left(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\right) = (1-q)^2$

1. In many places in the manuscript, we will use Landau's notations to compare functions. We recall here the meaning of the different notations we will use.

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two scalar functions. We have:

For $x \to x_0 \in \mathbb{R}$		
$f(x) = o(g(x))$	$\iff$	$\forall \epsilon > 0, \exists \delta > 0, \forall x \in \mathbb{R}, \lvert x - x_0\rvert < \delta \implies \lvert f(x)\rvert \leq \epsilon \lvert g(x)\rvert$
$f(x) = \mathcal{O}(g(x))$	$\iff$	$\exists M > 0, \exists \delta > 0, \forall x \in \mathbb{R}, \lvert x - x_0\rvert < \delta \implies \lvert f(x)\rvert \leq M \lvert g(x)\rvert$
$f(x) = \Theta(g(x))$	$\iff$	$\exists m, M > 0, \exists \delta > 0, \forall x \in \mathbb{R}, \lvert x - x_0\rvert < \delta \implies m\lvert g(x)\rvert \leq \lvert f(x)\rvert \leq M\lvert g(x)\rvert$
$f(x) = \omega(g(x))$	$\iff$	$\forall M > 0, \exists \delta > 0, \forall x \in \mathbb{R}, \lvert x - x_0\rvert < \delta \implies M\lvert g(x)\rvert \leq \lvert f(x)\rvert$
For $x \to +\infty$		
$f(x) = o(g(x))$	$\iff$	$\forall \epsilon > 0, \exists \xi \in \mathbb{R}, \forall x \in \mathbb{R}, x > \xi \implies \lvert f(x)\rvert \leq \epsilon \lvert g(x)\rvert$
$f(x) = \mathcal{O}(g(x))$	$\iff$	$\exists M > 0, \exists \xi \in \mathbb{R}, \forall x \in \mathbb{R}, x > \xi \implies \lvert f(x)\rvert \leq M \lvert g(x)\rvert$
$f(x) = \Theta(g(x))$	$\iff$	$\exists m, M > 0, \exists \xi \in \mathbb{R}, \forall x \in \mathbb{R}, x > \xi \implies m\lvert g(x)\rvert \leq \lvert f(x)\rvert \leq M\lvert g(x)\rvert$
$f(x) = \omega(g(x))$	$\iff$	$\forall M > 0, \exists \xi \in \mathbb{R}, \forall x \in \mathbb{R}, x > \xi \implies M\lvert g(x)\rvert \leq \lvert f(x)\rvert$

1.2.3 Statistical estimation

In this section, we present the most classical estimators of the measures of association introduced in the previous section, including the linear correlation due to its widespread use, even if it is not a measure of association.

The sampling definition of the linear correlation coefficient is given by:

Definition 1.37. Let $\left((X_1^k, X_2^k)\right)_{k=1,\dots,N}$ be a sample of size N of the random vector $\boldsymbol{X} = (X_1, X_2)$. The sampling linear correlation coefficient $\hat{\rho}_N(X_1, X_2)$ is defined by

$$\hat{\rho}_N(\boldsymbol{X}) = \frac{\sum_{k=1}^n (X_1^k - \bar{X}_1)(X_2^k - \bar{X}_2)}{\sqrt{\sum_{k=1}^N (X_1^k - \bar{X}_1)^2 \sum_{k=1}^N (X_2^k - \bar{X}_2)^2}} \tag{1.31}$$

where $\bar{X}_1 = \frac{1}{N}\sum_{k=1}^N X_1^k$ and $\bar{X}_2 = \frac{1}{N}\sum_{k=1}^N X_2^k$.

The asymptotic properties of this estimator are given in the following theorems [Gay51, Equations 53 and 54]:

Theorem 1.38. *Let $\boldsymbol{X}$ be a bi-dimensional random vector with finite second moments $\mathbb{E}[X_1^2] < \infty$ and $\mathbb{E}[X_2^2] < \infty$. Then:*

$$\hat{\rho}_N(\boldsymbol{X}) \overset{a.s}{\to} \rho(\boldsymbol{X}) \text{ when } N \to \infty$$

Theorem 1.39. *Let $\boldsymbol{X}$ be a bi-dimensional random vector with finite fourth-order moments $\mathbb{E}[X_1^4] < \infty$ and $\mathbb{E}[X_2^4] < \infty$. Then:*

$$\sqrt{N}\left(\hat{\rho}_N(\boldsymbol{X}) - \rho(\boldsymbol{X})\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \sigma_\rho^2) \text{ when } N \to \infty$$

where the asymptotic variance σ_ρ^2 is given by:

$$\sigma_\rho^2 = \left(1 + \frac{\rho^2(\boldsymbol{X})}{2}\right)\frac{m_{22}}{m_{20}m_{02}} + \frac{\rho^2(\boldsymbol{X})}{4}\left(\frac{m_{40}}{m_{20}^2} + \frac{m_{04}}{m_{02}^2} - \frac{4}{m_{11}}\left(\frac{m_{31}}{m_{20}} + \frac{m_{13}}{m_{02}}\right)\right)$$

where $m_{k\ell} = \mathbb{E}\left[(X_1 - \mu_1)^k (X_2 - \mu_2)^\ell\right]$, $\mu_1 = \mathbb{E}[X_1]$ and $\mu_2 = \mathbb{E}[X_2]$.

The notion of rank plays a key role in the estimation of measures of association.

Definition 1.40. Let $(X^k)_{k=1,\dots,N}$ be a sample of size N of the random variable X and $\sigma \in \mathfrak{S}_N$ a random permutation such that $X_{\sigma(1)} \leq \dots \leq X_{\sigma(N)}$ a.s. (such a permutation is almost surely unique if X is continuous). The rank of X^k is defined by:

$$\text{rank}(X^k) = \sigma^{-1}(k)$$

It is the random position of X^k in the sorted sample $(X_{\sigma(k)})_{k=1,\dots,N}$.

The definition of the Spearman rho coupled with the expression of the linear correlation coefficient estimator given previously, we estimate the Spearman rho as being the linear correlation coefficient of the ranks of the observations. For the case where there is no tie in the observations, which is the case of interest for applications with continuous distributions, we are able to express this estimator in a more compact way:

Definition 1.41. Let $\left((X_1^k, X_2^k)\right)_{k=1,\dots,N}$ be a sample of size N of the random vector $\boldsymbol{X} = (X_1, X_2)$. The Spearman rho estimator $\hat{\rho}_{S,N}(\boldsymbol{X})$ is the linear correlation coefficient estimator applied to the ranks of the given sample:

$$\hat{\rho}_{S,N}(\boldsymbol{X}) = \frac{\sum_{k=1}^{n}\left(\mathrm{rank}(X_1^k) - \overline{\mathrm{rank}}(X_1)\right)\left(\mathrm{rank}(X_2^k) - \overline{\mathrm{rank}}(X_2)\right)}{\sqrt{\sum_{k=1}^{N}\left(\mathrm{rank}(X_1^k) - \overline{\mathrm{rank}}(X_1)\right)^2 \sum_{k=1}^{N}\left(\mathrm{rank}(X_2^k) - \overline{\mathrm{rank}}(X_2)\right)^2}} \tag{1.32}$$

where $\overline{\mathrm{rank}}(X_1) = \frac{1}{N}\sum_{k=1}^{N}\mathrm{rank}(X_1^k)$ and $\overline{\mathrm{rank}}(X_2) = \frac{1}{N}\sum_{k=1}^{N}\mathrm{rank}(X_2^k)$. If there is no tie, i.e. $\forall i, j,\ (i \neq j) \Rightarrow (X_1^i \neq X_1^j \text{ or } X_2^i \neq X_2^j)$, the sampling Spearman rho $\hat{\rho}_{S,N}(X_1, X_2)$ is given by

$$\hat{\rho}_{S,N}(\boldsymbol{X}) = 1 - \frac{6\sum_{k=1}^{N}\left(\mathrm{rank}(X_1^k) - \mathrm{rank}(X_2^k)\right)^2}{N(N^2-1)} \tag{1.33}$$

The asymptotic properties of this estimator are given in the following theorems, deduced from the corresponding theorems for the linear correlation coefficient and the fact that $F_i(X_i)$ $(i = 1, 2)$ is uniformly distributed over $[0, 1]$ for continuous F_i:

Theorem 1.42. *Let $\boldsymbol{X}$ be a bi-dimensional continuous random vector. Then:*

$$\hat{\rho}_{S,N}(\boldsymbol{X}) \xrightarrow{a.s} \rho_S(\boldsymbol{X}) \text{ when } N \to \infty$$

where $\rho_S(\boldsymbol{X})$ is the Spearman rho between X_1 and X_2, as defined in Definition 1.2.2.

Theorem 1.43. *Let $\boldsymbol{X}$ be a bi-dimensional continuous random vector. Then:*

$$\sqrt{N}\left(\hat{\rho}_{S,N}(\boldsymbol{X}) - \rho_S(\boldsymbol{X})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{\rho_S}^2) \text{ when } N \to \infty$$

where the asymptotic variance $\sigma_{\rho_S}^2$ is given by:

$$\sigma_{\rho_S}^2 = \left(1 + \frac{\rho_S(\boldsymbol{X})^2}{2}\right)\frac{4(5 + 192\eta_{10})}{3(4\eta_{00}-1)^2} + \frac{\rho_S(\boldsymbol{X})^2}{4}\left(\frac{342}{125} - \frac{12}{5}\left(\frac{24(\eta_{20}+\eta_{02}) - 1)}{4\eta_{00}-1}\right)\right)$$

where $\eta_{k\ell} = \iint_{[0,1]^2}\left(u_1 - \frac{1}{2}\right)^k\left(u_2 - \frac{1}{2}\right)^\ell C(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2$ *and C is the copula of $\boldsymbol{X}$.*

The definition of the Kendall tau leads to an estimator that can also be expressed easily in terms of the discordance and concordance of the observations when there is no tie. In this case, the estimator reads:

Definition 1.44. Let $\left((X_1^k, X_2^k)\right)_{k=1,\dots,N}$ be a sample of size N of the random vector $\boldsymbol{X} = (X_1, X_2)$. The sampling Kendall tau $\hat{\tau}_N(X_1, X_2)$ is given by

$$\hat{\tau}_N(\boldsymbol{X}) = \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\mathrm{sgn}(X_1^i - X_1^j)\,\mathrm{sgn}(X_2^i - X_2^j) \tag{1.34}$$

The asymptotic properties of this estimator are given in the following theorems:

Theorem 1.45. *Let $\boldsymbol{X}$ be a bi-dimensional random vector. Then:*

$$\hat{\tau}_N(\boldsymbol{X}) \overset{a.s}{\to} \tau(\boldsymbol{X}) \text{ when } N \to \infty$$

Theorem 1.46. *Let $\boldsymbol{X}$ be a bi-dimensional random vector. Then:*

$$\sqrt{N}\left(\hat{\tau}_N(\boldsymbol{X}) - \tau(\boldsymbol{X})\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \sigma_\tau^2) \text{ when } N \to \infty$$

where the asymptotic variance σ_τ^2 is given by:

$$\sigma_\tau^2 = 4\,\mathbf{Var}\left[\mathbb{E}\left[\operatorname{sgn}(X_1 - X_1')\operatorname{sgn}(X_2 - X_2') \,|\, X_1, X_2\right]\right]$$

where $\boldsymbol{X}' = (X_1', X_2')$ is an independent copy of $\boldsymbol{X}$.

In contrast with the previous measures, no estimator for the upper or lower tail dependence coefficients has become standard, despite the large amount of research in this area, in relation with the estimation of **extrem values copulas** (see [KN00]). Being defined as a limit, these quantities are difficult to estimate, and except in fully parametric contexts, there will always be a trade-off between the bias (taking into account a large amount of the available data, including non-extreme ones) and the variance (taking into account only the most extreme data) of the estimator. We restrict the presentation to non-parametric estimators, based on the **empirical copula** defined here:

Definition 1.47. Let $\left((X_1^k, X_2^k)\right)_{k=1,\dots,N}$ be a sample of size N of the random vector $\boldsymbol{X} = (X_1, X_2)$. The **empirical copula** $\hat{C}_N$ of this sample is the bivariate function defined by:

$$\forall (u_1, u_2) \in [0,1]^2, \quad \hat{C}_N(u_1, u_2) = \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{(\operatorname{rank}(X_1^k) \le N u_1,\, \operatorname{rank}(X_2^k) \le N u_2)} \tag{1.35}$$

We present a non-parametric estimators of the upper-tail coefficient based on the empirical copula of block maxima proposed in [SS04] and in [FJS05]:

Definition 1.48. Let $\left((X_1^k, X_2^k)\right)_{k=1,\dots,N}$ be a sample of size N of the random vector $\boldsymbol{X} = (X_1, X_2)$. Let m be a positive integer and $\ell = [N/m]$. We consider the sample $\left((x_1^{*j}, x_2^{*j})\right)_{j=1,\dots,m}$ of **componentwise block maxima**:

$$x_1^{*j} = \max\left\{X_1^i, i = 1 + (j-1)\ell, \dots, j\ell\right\}$$
$$x_2^{*j} = \max\left\{X_2^i, i = 1 + (j-1)\ell, \dots, j\ell\right\}$$

for $j = 1, \dots, m$. For a given integer threshold $0 < k(m) < m$, the upper tail coefficient λ_U can be estimated by:

$$\hat{\lambda}_{U,m}(\boldsymbol{X}) = 2 - \frac{1 - \hat{C}_m\left(\frac{m-k}{m}, \frac{m-k}{m}\right)}{1 - \frac{m-k}{m}}$$

The parameters m and k allow to deal with the bias/variance trade-off. The properties of this estimator are given in the following theorems, given in [SS04, Theorem 7] and [SS04, Corollary 2]:

Theorem 1.49. *Let $\boldsymbol{X}$ be a bi-dimensional random vector with continuous marginal distribution function. If the upper tail copula $\Lambda_U \neq 0$ exists and $k(m)$ is such that $k(m)/\log\log m \to 0$ as $m \to \infty$. Then:*

$$\hat{\lambda}_{U,m}(\boldsymbol{X}) \overset{a.s}{\to} \lambda_U(\boldsymbol{X}) \text{ when } m \to \infty$$

Theorem 1.50. *Let $\boldsymbol{X}$ be a bi-dimensional random vector with continuous marginal distribution function. If the upper tail copula $\Lambda_U \neq 0$ exists, possesses continuous partial derivatives, and satisfies the additional second order condition: it exists a function $A : \mathbb{R}_+ \to \mathbb{R}_+$ such that $A(t) \to 0$ as $t \to \infty$ and:*

$$\lim_{t\to\infty} \frac{\Lambda_U(\boldsymbol{u}) - (1-t)C(1-u_1/t, 1-u_2/t)}{A(t)} = g(\boldsymbol{u}) < \infty$$

locally uniformly for $\boldsymbol{u} \in [0,1]^2$ and some nonconstant function g.

Then, if $\sqrt{k(m)}A(m/k(m)) \to 0$ as $m \to \infty$:

$$\sqrt{k(m)}\left(\hat{\lambda}_{U,m}(\boldsymbol{X}) - \lambda_U(\boldsymbol{X})\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \sigma_U^2) \text{ when } m \to \infty$$

with

$$\begin{aligned}\sigma_U^2 =& \lambda_U(\boldsymbol{X}) + \left(\frac{\partial}{\partial x}\Lambda_U(1,1)\right)^2 + \left(\frac{\partial}{\partial y}\Lambda_U(1,1)\right)^2 + \\ & 2\lambda_U(\boldsymbol{X})\left(\left(\frac{\partial}{\partial x}\Lambda_U(1,1) - 1\right)\left(\frac{\partial}{\partial y}\Lambda_U(1,1) - 1\right) - 1\right).\end{aligned}$$

Conclusion

In this introductory chapter, we have introduced several concepts and measures linked with dependence modeling that will be used in the sequel of the manuscript. It covers both the probabilistic aspects linked with the distribution function of a random vector and the dependence quantification through scalar measures. We have also given some elements on statistical estimation of these measures given a set of multidimensional data.

Chapter 2

A practical approach to dependence modeling using copulas

From the previous chapter, the copula concept appears as the natural way to express the dependence structure of a random vector. Nevertheless, facing a problem of probabilistic uncertainty propagation, one can hope that some limited but strategically chosen information on dependence may lead to the description of a dependence structure surely partial but actually enough to compute the decision criteria with enough accuracy.

In this chapter, we formalize and explore numerically the implicit assumption that a set of measures of association is able to capture enough dependence information to provide a meaningful estimation of a probability of failure for decision making under uncertainty.

This work has been published in [DL09].

2.1 On the good usage of measures of association for probability estimation

To formalize these considerations, in this section, we introduce the notion of **dependence information**, that denotes the available information regarding the dependence structure, and the notion of ε**-synthesis** that quantifies the pertinence of the dependence information for the evaluation of a specific decision criterion. We will focus our attention on a criterion based on the probability of failure as defined in the introduction.

2.1.1 Dependence information

The formal definition of **dependence information** and ε**-synthesis** of this information are as follows.

Definition 2.1. Let $\mu_1, \ldots, \mu_k$ be k measures of association and $[a_1, b_1], \ldots, [a_k, b_k]$ be their possible range. We call **dependence information** the set $I = \{(\mu_i, [a_i, b_i])\}_{i=1,\ldots,k}$.

Remark 2.2. We can have $a_i = b_i$, which means that the value of the i-th measure of association μ_i is known. We can as well have $a_i = -\infty$ or $b_i = +\infty$, which means that one can only give an upper or a lower bound on the value of μ_i.

Definition 2.3. Let C be a copula and I a dependence information. I and C are said to be **compatible** if and only if :

$$\forall i \in \{1, \dots, k\}, \quad \mu_i(C) \in [a_i, b_i] \tag{2.1}$$

otherwise I and C are said to be **incompatible**.
We note:

$$\mathcal{E}_{comp}(I) = \{C \,|\, C \text{ is a copula compatible with } I\}$$

the set of copulas compatible with I.

Remark 2.4. It might happen that $\mathcal{E}_{comp}(I) = \emptyset$. For example, if we take $k = 2$, $\mu_1 = \rho_S$, $\mu_2 = \tau$, $a_1 = b_1 = -1$ and $a_2 = b_2 = 1$ then there exists no copula C such that I and C are compatible, because for every copula C the following bounds hold (see [Nel06]):

$$-1 \leq 3\tau(C) - 2\rho_S(C) \leq 1 \tag{2.2}$$

Definition 2.5. Let $\boldsymbol{X}$ be a random vector with known marginal distributions and its copula C_X, partially known through the dependence information I. Let g be a numerical model and s a scalar threshold. We define the bounds P_{min}, P_{max} and the ratio r as:

$$P_{min} = \min_{C \in \mathcal{E}_{comp}(I)} P(C), \quad P_{max} = \max_{C \in \mathcal{E}_{comp}(I)} P(C) \text{ and } r = P_{max}/P_{min} \tag{2.3}$$

where $P(C) = \mathbb{P}\left(g(\boldsymbol{X'}) > s\right)$ with $\boldsymbol{X'}$ being a random vector with copula C and the same marginal distributions as $\boldsymbol{X}$.

We say that I is a ε-synthesis with respect to $\mathbb{P}\left(g(\boldsymbol{X}) \geq s\right)$ if and only if $\varepsilon \geq r$.

Remark 2.6. We have obviously $P_{min} \leq P(C_X) \leq P_{max}$ because $C_X \in \mathcal{E}_{comp}(I)$. Furthermore, it is possible that $P_{min} = P_{max} = 0$, in which case the ε-synthesis is not defined. In this case, we adopt the convention that $r = +\infty$ and I is then a ∞-synthesis.

The notion of ε-synthesis allows to quantify the maximum possible dispersion of the probability of failure when the dependence structure varies whithin the set of copulas compatible with the given dependence information.

As when one performs an effective computation of such a probability, one has to choose a specific copula C_0 based on the available dependence information, it is of interest to quantify the **dispersion** between the value obtained using C_0 instead of the unknown copula C_X:

Definition 2.7. Let I be a dependence information and C_0 a copula compatible with I, said to be a **reference copula**. The **dispersion** $r(C_0)$ with respect to C_0 is defined as:

$$r(C_0) = \max_{C \in \mathcal{E}_{comp}(I)} \left(P_{max}/P(C_0), P(C_0)/P_{min}\right) \tag{2.4}$$

where P_{min} and P_{max} have been defined in (2.3).

The optimal choice for C_0 is to take a copula that minimizes $r(C_0)$. It is a hard task in the general case, so the minimization is done on a reduced set of copulas, typically by inspection in a finite set of copulas.

2.1.2 Usefulness of a dependence information

Depending on the usage that will be made of the value of $P(C)$, one can be interested by a more or less precise evaluation of $P(C)$. For example, to determine in which class of risk a specific industrial installation is, one is interested in the evaluation of the order of magnitude of $P(C)$ and not its precise value. Thus, we define two arbitrary levels of ε-synthesis that will be adapted to different scenarii of exploitation of P: a **qualitative** ε-synthesis and a **quantitative** ε-synthesis.

Definition 2.8. Let I be a ε-synthesis. We say that the dependence information I is **quantitative** if $1 \leq \varepsilon \leq 1.5$, and I is **qualitative** if $1.5 \leq \varepsilon \leq 10$.

In order to illustrate the interest of a qualitative estimation of a probability of failure, Table 2.1 gives the scale of risk defined in [RTC96], a technical report used by the Federal Aviation Administration (FAA) to evaluate the risk induced by any disfunction per hour of flight on an aircraft.

Risk	Probability level	Frequency
catastophic	$P(C) < 10^{-9}$	extremely improbable
hazardous	$10^{-9} < P(C) < 10^{-7}$	extremely remote
major	$10^{-7} < P(C) < 10^{-5}$	remote
minor	$10^{-5} < P(C) < 10^{-3}$	reasonably probable
minor	$10^{-3} < P(C) < 10^{-2}$	probable

Table 2.1: Risk classes according to DO-233

We see that an estimation precise whithin an order of magnitude of the probability of failure has still some interest within this context.

2.2 Numerical experiments

With the measures of association presented in the previous chapter, have we progressed in our quest of a compact but still accurate representation of the stochastic dependence? In order to give a more quantitative analysis of the role played by the copula in reliability analysis, we present here a series of numerical results related to a problem arising in the aeronautic industry.

2.2.1 Risk induced by the portable electronic devices in aeronautic

There is a strong demand from the customers of flight companies to have the opportunity to use their portable electronic devices (PEDs) during a flight: mobile phones, laptops and so on. Up to now, this usage is prohibited, due to the risk of interference between the PEDs and the flight control system.
This position results from an analysis made of two parts: first, physical measurements have been carried out for a wide variety of PEDs in order to identify their spectral emission (mainly an electromagnetical energy as a function of the frequency). Then, several configurations have been numerically studied, using a worst-case analysis : the most penalizing situation in terms of amplitude and frequency of the emission and location of the emitter is considered, regardless of its probability of occurence. It is obviously not satisfying, since we base a decision on a very penalizing and very unlikely situation. That is

why a reliablility approach of the problem has been performed. As the PED does not have to be certified in the range of frequencies the aeronautic industry is interested in, the two most important quantities that are the the main frequency of the PED (corresponding to its maximum emission) and the energy emitted at this frequency, are very uncertain and **dependent** quantities. We model these two uncertain parameters by a bi-dimensional random vector $\boldsymbol{X}$ and we explore numerically the effect of the dependence between these quantities.

2.2.2 Probabilistic approach for PED certification

Several classes of risk have been defined for qualifying the risk induced by PEDs. These classes are recalled in Table 2.1.

The effect of a coupling between the emission of a PED and an antenna of the aircraft might have a major impact on the safety of the flight. To perform the analysis, we will do the following steps: First, we choose a measure of association between the main frequency and the energy and we estimate it thanks to physical experiments. Then, we explore numerically the effect of the copula by choosing different copulas that leads to the same value for the selected measure of association on the probability $P(C)$ of coupling between the PED and the antenna.

We will not give more details on both the industrial context and the numerical model of coupling, but we will perform a parametric study on a simplified and more generic description of the problem.

2.2.3 Generic problem in the normalized space

The generic problem we consider is the following:

– The random vector $\boldsymbol{X}$ describing the uncertain physical parameters has a joint distribution function F_{12} such that:

$$\forall (x_1, x_2) \in \mathbb{R}^2, F_{12}(x_1, x_2) = C(\Phi(x_1), \Phi(x_2)) \tag{2.5}$$

where Φ is the distribution function of the standard normal distribution $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\,\mathrm{d}t$ and C is a copula chosen in Table 1.1, with two specific choices for the degree of freedom of the Student copulas: $\nu = 3$ and $\nu = 10$. This choice has been motivated by the fact that this set includes members that have any combination of tail dependence (null/positive lower tail dependence, null/positive upper tail dependence). For the sequel, the definition of ε-synthesis is restricted to this specific set of copulas.

– The physical model g is an affine function from $\mathbb{R}^2$ to $\mathbb{R}$ defined by:

$$\forall (x_1, x_2) \in \mathbb{R}^2, g_{\alpha,\beta}(x_1, x_2) = x_1 \cos\alpha + x_2 \sin\alpha - \beta \tag{2.6}$$

– The quantity of interest is $Y = g_{\alpha,\beta}(\boldsymbol{X})$, and the failure event $\{\omega \,|\, Y(\omega) \geq 0\}$.

The failure domain $F_{\alpha,\beta} = \{\boldsymbol{x} \in \mathbb{R}^2 \,|\, g_{\alpha,\beta}(x_1, x_2) \geq 0\}$ is a half-space, see Figure 2.1. The parameter β is the distance from the origin which allows to change the probability level of the problem, while α is the angle of rotation of the failure domain, it allows to change the influence of a given dependence structure, between $\alpha = 0$ (no influence of the copula since then $P(C) = \Phi(\beta)$) and $\alpha = \pi/4$ (maximum influence of the dependence

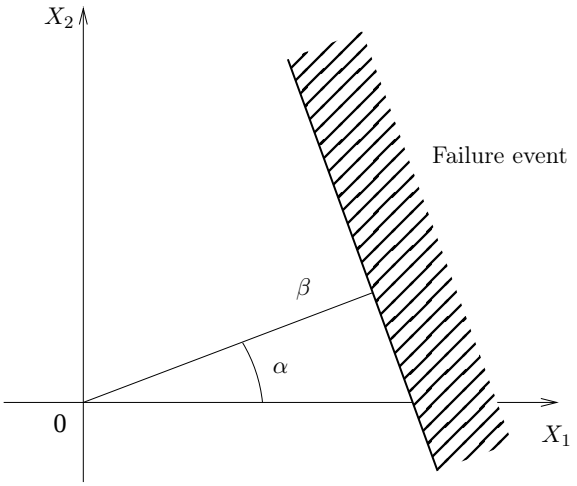


Figure 2.1: Generic failure domain in the normalized space.

structure, due to the symmetry of the chosen copulas).

The probability $\mathbb{P}(\boldsymbol{X} \in F_{\alpha,\beta})$ will be written $P_{\alpha,\beta}(C)$ in order to underline its dependence with respect to the copula C. It is computed using an adaptive Gauss-product integration that gives very accurate results in a reasonable amount of time. The motivation for using such an unusual integration procedure in the context of uncertainty propagation, instead of e.g. a Monte Carlo method, is that we are focused on the influence of the dependence modeling and we do not want to add further approximation induced by the numerical method used to compute the probability of failure. The adaptive Gaussian quadrature allows for a numerical integration with a relative accuracy better than 10^{-5}, so these values will be considered as exact for our purpose.

We will consider three kinds of dependence information for the numerical experiments, as presented in Table 2.2.

Experiment 1	I is restricted to the value of ρ_S.
Experiment 2	I is restricted to the value of ρ_S and the information that $\lambda_U > 0$(*).
Experiment 3	I is restricted to the value of λ_U.

Table 2.2: Three numerical experiments with different kinds of information on the dependence.
(*) we are not supposed here to know the value of λ_U.

Each numerical experiment is parametric on (α, β) in order to see both the influence of the coupling between the random variables due to the failure domain as well as the level of probability we are computing. We will monitor the following quantities:

– $r_{max}(\beta) = \max_{\alpha} r(\alpha, \beta)$ with

$$r(\alpha, \beta) = \frac{P_{max}(\alpha, \beta)}{P_{min}(\alpha, \beta)}$$

and using relation (2.3):

$$P_{min}(\alpha,\beta) = \min_{C\in\mathcal{E}_{comp}(I)} P_{\alpha,\beta}(C)$$
$$P_{max}(\alpha,\beta) = \max_{C\in\mathcal{E}_{comp}(I)} P_{\alpha,\beta}(C)$$

- $\alpha^* = \min\{\alpha \,|\, r(\alpha,\beta) = r_{max}(\beta)\}$, which is the smallest value of α such that $r(\alpha,\beta) = r_{max}(\beta)$;
- $P_{max}(\alpha^*,\beta)$ and $P_{min}(\alpha^*,\beta)$, which are the bounds of the interval of variation of $P_{\alpha^*,\beta}(C)$ when C is any copula compatible with the given dependence information.
- $r_{max}(C_0,\beta) = \max_{\alpha} r_{\alpha,\beta}(C_0)$, cf. (2.4);
- $\alpha^*(C_0) = \min\{\alpha \,|\, r_{\alpha,\beta}(C_0) = r_{max}(C_0,\beta)\}$, which is the smallest value of α such that $r_{\alpha,\beta}(C_0) = r_{max}(C_0,\beta)$;
- $P(C_0,\beta) = P_{\alpha^*(C_0),\beta}(C_0)$ which gives the evolution of the probability associated with the reference copula with respect to β.

We also show the bounds corresponding to a quantitative and qualitative ε-synthesis on the figures related to the evolution of $r_{max}(\beta)$ and $r_{max}(C_0,\beta)$ with respect to β (see Figures 2.3, 2.5 and 2.7).

2.2.4 First experiment: the value of the Spearman rho is given

In this experiment, we suppose that the copula C_X is such that the Spearman correlation ρ_S of $\boldsymbol{X}$ is equal to $\rho_S = 0.5$. The set of copulas compatible with this dependence information is given in Table 2.3.

Copula	Parameter.
Normal	$\rho = 0.518$
Gumbel	$\theta = 1.54$
Frank	$\theta = 3.45$
Clayton	$\theta = 1.08$
Comp. Clayton	$\theta = 1.08$
Student copula, with $\nu = 3$	$\rho = 0.537$
Student copula, with $\nu = 10$	$\rho = 0.523$

Table 2.3: Three numerical experiments with different kinds of information on the dependence.

Among these seven copulas, only the four last ones have positive upper tail dependence. We also show the results obtained when using the independent copula (called Normal indep. on Figure 2.2), in order to better see the influence of the dependence structure.

The extensive numerical exploration of the effect of the copula on a failure probability is summarized on Figure 2.2. We show the case corresponding to $\alpha = \pi/4$, as it is the situation where the results are the most spread amongst the different copulas. We distinguish 3 zones:

- The first zone (zone 1 on the figure) corresponds to failure domains for which the probability does not vary by more than a factor 1.5 when the copula changes. We say that we can have a **quantitative** estimate of the true probability when we take any copula available and fix its parameters such that the associated Spearman rho is equal to the needed value (here 0.5). This zone corresponds to probabilities of at least 0.1, which is by no way the level we are interested in.

– The second zone (zone 2 on the figure) corresponds to failure domains for which the probability varies by more than a factor 1.5, but by less than a factor 10 between the extremal values. We say that we can still have a **qualitative** estimate of the true probability under the same conditions than in the previous zone. For this zone, we are around probabilities of 0.005, which is still too high for our purpose.
– The third zone (zone 3 on the figure) corresponds to failure domains where the knowledge of the Sperman rho value is not enough to estimate the probability by at least one order of magnitude. It is precisely in this zone that we have to work.

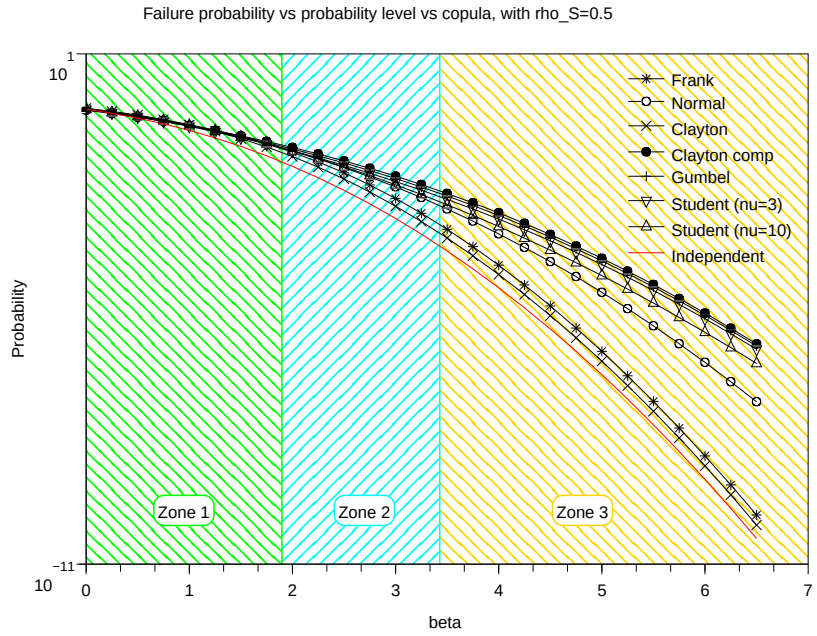


Figure 2.2: Evolution of the probability of failure with β, for different copulas with the same Spearman rho ($\rho_S = 0.5$). We highlight three regions according to the quality of the ε-synthesis of this information of dependence.

Using these graphs, we decide to take the normal copula as the reference copula C_0. The value of the different quantities we are monitoring are presented in Table 2.4.

β	$r_{max}(\beta)$	$P_{min}(\alpha^*, \beta)$	$P_{max}(\alpha^*, \beta)$	$\sqrt{P_{max}P_{min}}$	$r_{max}(C_0, \beta)$	$P(C_0, \beta)$
1.89	1.5	$6.5\,10^{-2}$	$8.7\,10^{-2}$	$7.5\,10^{-2}$	–	–
2.18	–	–	–	–	1.5	$5.2\,10^{-2}$
3.41	10.0	$1.1\,10^{-3}$	$8.6\,10^{-3}$	$3.1\,10^{-3}$	–	–
4.07	–	–	–	–	10.0	$5.8\,10^{-4}$
6.5	$2.3\,10^{4}$	$8.3\,10^{-11}$	$1.9\,10^{-6}$	$1.3\,10^{-8}$	$7.9\,10^{2}$	$6.6\,10^{-8}$

Table 2.4: Maximum variation of $P_{\alpha,\beta}(C)$ and maximum dispersion with respect to the normal copula $P_{\alpha,\beta}(C_0)$ when $\rho_S = 0.5$ and $\alpha \in [0, \pi/4]$, for some characteristic values of β.

On Figure 2.3, we see the evolution of $r_{max}(\beta)$ and $r_{max}(C_0, \beta)$ that shows the transition between the quantitative and the qualitative estimate, with and without a reference

copula.

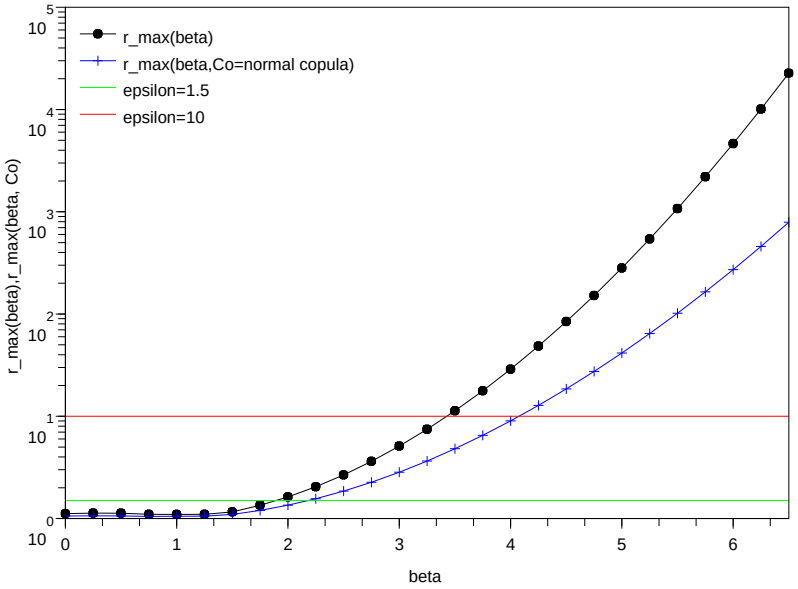


Figure 2.3: Evolution of $r_{max}(\beta)$ and $r_{max}(C_0, \beta)$ for $\rho_S = 0.5$. We see that the range over which the estimate is quantitative is quite limited.

2.2.5 Second experiment: the value of the Spearman rho is given, and we know that a positive upper tail dependence exists

In this second experiment, we suppose that the copula C_X is such that the Spearman correlation ρ_S of $\boldsymbol{X}$ is equal to $\rho_S = 0.5$ and that the coefficient of upper tail dependence exists and is positive $\lambda_U > 0$. The copulas compatible with this dependence information are the Gumbel one, the complementary Clayton one and both Student ones, parameterized as previously.

The extensive numerical exploration of the effect of the copula on a failure probability is summarized on Figure 2.4, for $\alpha = \pi/4$. The two zones have the same meaning as in the first experiment. We see that the additional dependence information of the existence of a positive tail dependence allows to have an estimate that is at least qualitative in the whole range of reliability index, and the range over which the estimate is quantitative is much larger than in the previous experiment.

Using these graphs, we decide to take the Student copula with $\nu = 3$ as the reference copula C_0. The value of the different quantities we are monitoring are presented in Table 2.5.

On Figure 2.5, we see the evolution of $r_{max}(\beta)$ and $r_{max}(C_0, \beta)$.

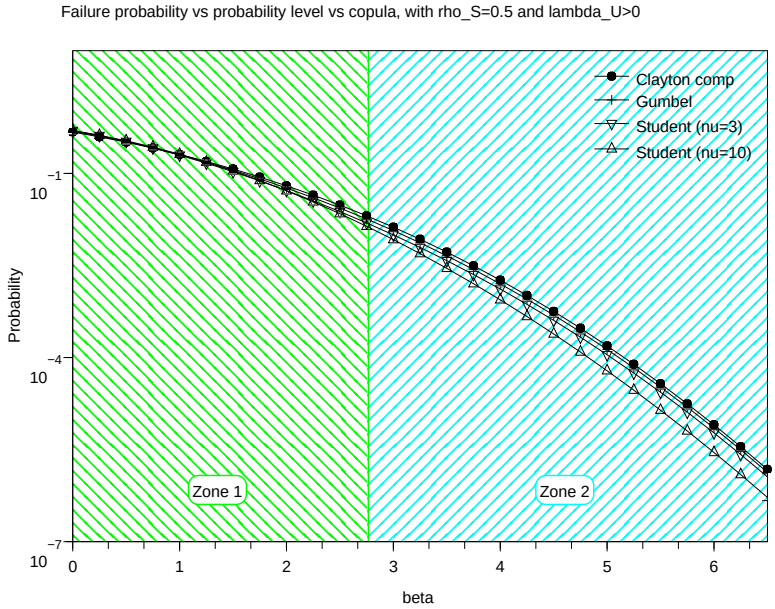


Figure 2.4: Evolution of the probability of failure with β, for different copulas with the same Spearman rho ($\rho_S = 0.5$) and positive coefficient of upper tail dependence ($\lambda_U > 0$). We highlight two regions according to the quality of the ε-synthesis of this information of dependence: in this case, the estimate is always at least qualitative.

β	$r_{max}(\beta)$	$P_{min}(\alpha^*, \beta)$	$P_{max}(\alpha^*, \beta)$	$\sqrt{P_{max}P_{min}}$	$r_{max}(C_0, \beta)$	$P(C_0, \beta)$
2.77	1.5	$1.4\,10^{-2}$	$2.0\,10^{-2}$	$1.7\,10^{-2}$	–	–
3.97	–	–	–	–	1.5	$2.2\,10^{-3}$
6.5	3.6	$5.2\,10^{-7}$	$1.9\,10^{-6}$	$9.9\,10^{-7}$	2.4	$1.2\,10{-}6$

Table 2.5: Maximum variation of $P_{\alpha,\beta}(C)$ and maximum dispersion with respect to the normal copula $P_{\alpha,\beta}(C_0)$ when $\rho_S = 0.5$, $\lambda_U > 0$ and $\alpha \in [0, \pi/4]$, for some characteristic values of β.

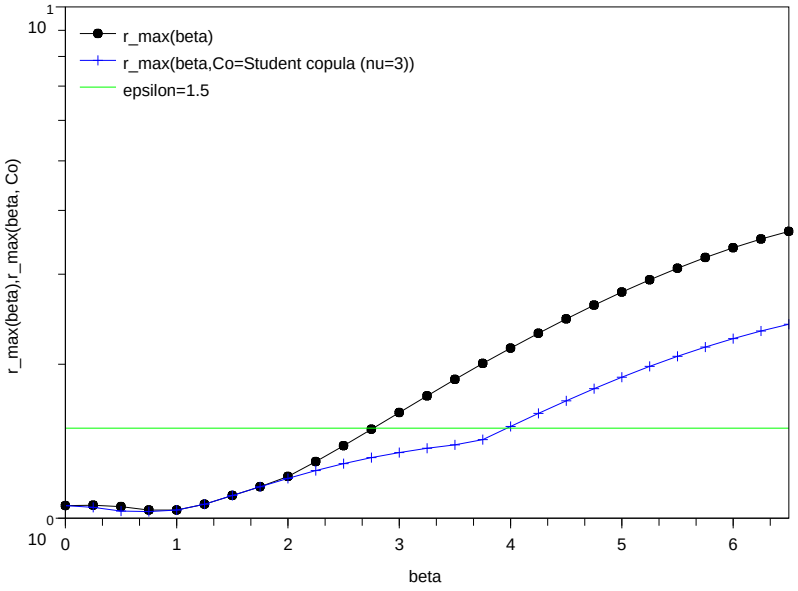


Figure 2.5: Evolution of $r_{max}(\beta)$ and $r_{max}(C_0, \beta)$ for $\rho_S = 0.5$. We see that the range over which the estimate is quantitative is larger, even if it is still quite limited, and we see that the estimate is always at least qualitative.

2.2.6 Third experiment: the value of the coefficient of upper tail dependence is known

In this last experiment, we suppose that the copula C_X has a known positive upper tail coefficient equal to $\lambda_U = 0.5$. The set of copulas compatible with this dependence information are given in Table 2.6 with their respective parameters.

Copula	Parameter.
Gumbel	$\theta = 1.71$
Comp. Clayton	$\theta = 1.00$
Student copula, with $\nu = 3$	$\rho = 0.759$
Student copula, with $\nu = 10$	$\rho = 0.915$

Table 2.6: Copula of the third numerical experiment, parameterized in order to verify $\lambda_U = 0.5$.

- The Gumbel copula, with $\theta = 1.71$,
- The complementary Clayton copula, with $\theta = 1.00$,
- The Student copula, with $\nu = 3$ and $\rho = 0.759$,
- The Student copula, with $\nu = 10$ and $\rho = 0.915$.

The extensive numerical exploration of the effect of the copula on a failure probability is summarized on Figure 2.6, for $\alpha = \pi/4$. The zone has the same meaning as in the first experiment. We see that the existence of a positive tail dependence and the knowledge of its value allows to have an estimate that is quantitative in the whole range of reliability

index.

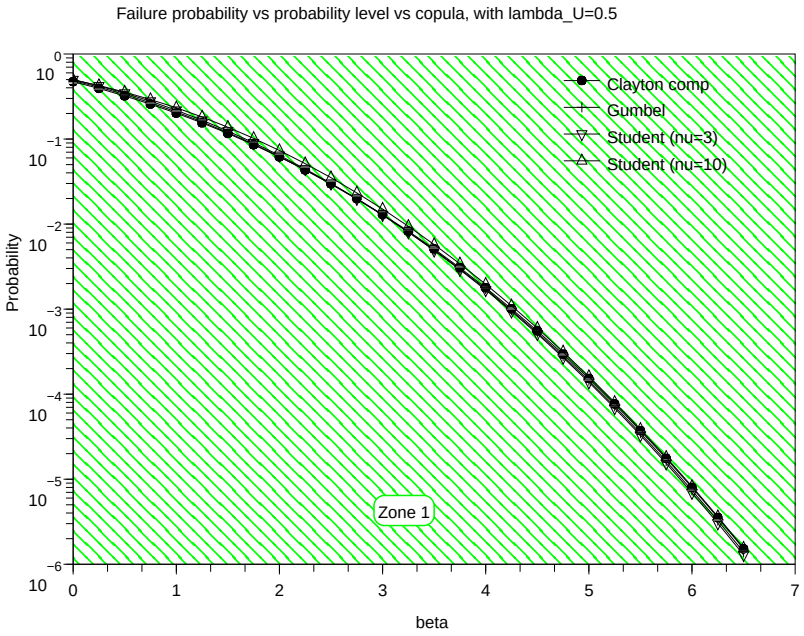


Figure 2.6: Evolution of the probability of failure with β, for different copulas with the same Spearman rho ($\rho_S = 0.5$) and positive coefficient of upper tail dependence ($\lambda_U > 0$). We see that the estimate is always quantitative.

Using these graphs, we decide to take the Student copula with $\nu = 3$ as the reference copula C_0. The value of the different quantities we are monitoring are presented in Table 2.7.

$\beta = 6.5$	$r_{max}(\beta)$	$P_{min}(\alpha^*, \beta)$	$P_{max}(\alpha^*, \beta)$	$\sqrt{P_{max}P_{min}}$	$r_{max}(C_0, \beta)$	$P(C_0, \beta)$
6.5	1.2	$1.3\,10^{-6}$	$1.5\,10^{-6}$	$1.4\,10^{-6}$	1.2	1.3 10−6

Table 2.7: Maximum variation of $P_{\alpha,\beta}(C)$ and maximum dispersion with respect to the normal copula $P_{\alpha,\beta}(C_0)$ when $\lambda_U = 0.5$ and $\alpha \in [0, \pi/4]$, for $\beta = 6.5$, which is the most dispersed case.

On Figure 2.7, we see the evolution of $r_{max}(\beta)$ and $r_{max}(C_0, \beta)$.

2.3 Conclusions

In this chapter, we have emphasized that taking into account the stochastic dependence structure is of uttermost importance for a correct evaluation of a probability of failure, and that this dependence cannot be properly represented by linear correlations as it is frequently done.

We have shown in the previous chapter that the correct way to **fully** represent the dependence is to determine the copula of the random vector. As this task can be uneasy, we reviewed several scalar measures of association that are more adapted than the linear

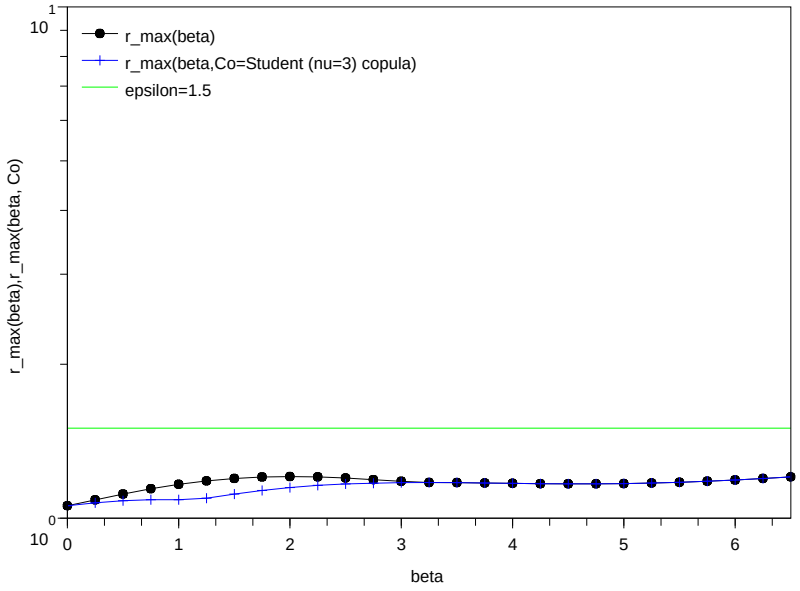


Figure 2.7: Evolution of $r_{max}(\beta)$ and $r_{max}(C_0, \beta)$ for $\lambda_U = 0.5$. We see that the estimate is quantitative on the whole range of β.

correlation to summarize the dependence structure. In this chapter, we have extended the notion of measure of association into a notion of dependence information, associated with a measure of the ability of such information to synthetize the dependence structure at a given level of precision.

We illustrated these concepts through a set of numerical experiments related to an aeronautical application. These experiments confirm that, in general, a single measure of association does not properly summarize the information related to the dependence structure, at least for the very low levels of probability occuring in reliability studies. With more dependence information, for example the knowledge of the value taken by another measure of association, it becomes possible to have a partial description of the dependence structure that seems to be enough to quantify even low levels of probability with an acceptable precision.

As the numerical experiments are based on very generic situations, we hope that these conclusions are of wider scope than the specific application we were interested in. Using only a small set of copulas, we were not able to fully explore the concept of dependence information, but we hope to have initiated the formalization of a usefull field of investigation.

Chapter 3

The usual Nataf transformation and copulas

As recalled in the general introduction, in the context of uncertainty management, one is interested in the evaluation of the probability of failure of a complex system for a given scenario of use. In many situations, the event associated with the failure of the system reduces to the exceedance of a given threshold s for a specific key characteristic Y of the system. For a continuous random vector of uncertain parameters $\boldsymbol{X}$ with joint density $p_{\boldsymbol{X}}$ linked to the quantity of interest though a complex numerical relation $Y = g(\boldsymbol{X})$ the characteristic variable of interest, the probability of failure we are interested in writes:

$$p = \mathbb{P}\left(Y \geq s\right) = \int_{\mathcal{D}_s} p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{3.1}$$

where $\mathcal{D}_s = \{\boldsymbol{x} \in \mathbb{R}^n / g(\boldsymbol{x}) \geq s\}$ is called the **failure domain**.

In the reliability context, authors such as in [DM05] mention two main difficulties:

- Neither g nor the boundary of $\mathcal{D}_s$ have explicite analytical expressions but they are typically given by a finite element model often requiring high CPU costs
- The joint density $p_{\boldsymbol{X}}$ is unknown.

The first point prevents from symbolic manipulations and the high CPU costs prevent from the use of classical numerical methods to evaluate integrals (Monte Carlo simulations, ...), whereas the second point raises the problem of modeling a joint probability distribution based only on information often reduced to the marginal distributions of $\boldsymbol{X}$ and some linear correlation coefficients when one wants to take into account some dependence between the input parameters.

That is why many authors recommend the use of the Nataf isoprobabilistic transformation (see [DKL86b] and [Nat62]) to map the **physical space** of the probabilistic input data into the **standard space**, where all the variables are independent and follow the same normal distribution with zero mean and unit variance. Then, within the standard space, it is possible to make a first-order or second-order geometrical approximation of the boundary of the failure domain, which allows us to compute an approximation of p thanks to an analytic expression.

This method, widely used in probabilistic uncertainty propagation studies, gives to the Nataf transformation a key role in the probabilistic modeling of $\boldsymbol{X}$. This role is largely overlooked in most of the presentations of the transformation (see [DKL86b], [DM05]), the accent being made on its analytical properties. The main drawback of such presentations is that the practitioner is not informed of the probabilistic hypotheses he made by using

this transformation, hypotheses that can have a very significant impact on the result of the uncertainty propagation.

In this chapter, we rewrite the Nataf transformation thanks to the copula theory. This innovating point of view highlights the hypotheses necessary for the Nataf transformation, which makes it possible to understand plainly the limitations of its use.
In the first part of the chapter, we detail the Nataf transformation in its usual presentation (as found in e.g. [HL74]) and how it is usually used in probabilistic propagation of uncertainties (see e.g. [DKL86a]). We recall the interest of such a transformation and the related probabilistic indicators obtained as by-products.
The concept of copula gives a new insight on the isoprobabilistic Nataf transformation and its hypotheses : in the second part of the chapter, we demonstrate that the Nataf transformation makes the important hypotheses of a normal dependence structure for the random input vector $\boldsymbol{X}$ and maps it into a Gaussian vector with independent, zero mean and unit variance components.
Finally, we list all the hypotheses underlying the Nataf transformation and the possible risks associated with its use. In particular, we explain the probabilistic consequences of using a normal dependence structure and the difficulties related to its parameterization with a linear correlation matrix. This work has been published in [LD09c].

3.1 Traditional use of the Nataf transformation

Very often, probabilistic data available about the random vector $\boldsymbol{X}$ are the marginal distributions (which are supposed here to have finite second-order moments) with marginal distribution functions $F_1, \dots, F_n$ and, in the particular case of correlated components, the linear correlation matrix $\boldsymbol{R} = (r_{ij})_{ij}$ of $\boldsymbol{X}$. As a result of Theorem 1.30, we know that these two pieces of information are subject to compatibility conditions: some correlation matrices are impossible to obtain as the correlation matrix of **any** joint distribution with marginal distributions F_i, or with these marginal distribution functions and a specific copula. If we suppose that no such incompatibility occurs, we have

$$r_{ij} = \mathbb{E}\left[\left(\frac{X_i - \mu_i}{\sigma_i}\right)\left(\frac{X_j - \mu_j}{\sigma_j}\right)\right] \tag{3.2}$$

where μ_i and σ_i are the mean and standard deviation of X_i. We suppose that $\boldsymbol{R}$ is non-singular.

In order to perform reliability analysis such as the computation of a probability of failure, the probabilistic modeling is completed thanks to the Nataf transformation. From a purely analytic point of view, this transformation reads:

Definition 3.1. Let $F_1, \dots, F_n$ be n continuous univariate distribution functions and $\boldsymbol{R}_0$ be a definite positive matrix with unit diagonal. The Nataf transformation T is the composition of two functions $T = T_2 \circ T_1$ such that

$$\begin{array}{rl|ccc} T_1 : & \mathbb{R}^n & \to & \mathbb{R}^n \\ & \boldsymbol{x} & \mapsto & \boldsymbol{w} = \begin{pmatrix} \Phi^{-1} \circ F_1(x_1) \\ \vdots \\ \Phi^{-1} \circ F_n(x_n) \end{pmatrix} \end{array}$$

$$\begin{array}{rl|ccc} T_2 : & \mathbb{R}^n & \to & \mathbb{R}^n \\ & \boldsymbol{w} & \mapsto & \boldsymbol{u} = \boldsymbol{\Gamma}\,\boldsymbol{w} \end{array} \tag{3.3}$$

The matrix $\boldsymbol{\Gamma}$ is any matrix such that $\boldsymbol{\Gamma}^t\boldsymbol{\Gamma} = \boldsymbol{R}_0^{-1}$ or equivalently $\boldsymbol{R}_0 = \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\Gamma}^{-1}\right)^t$, and Φ is the distribution function of the univariate standard normal distribution defined in the previous chapter, see (2.5).

Remark 3.2. A common choice for $\boldsymbol{\Gamma}$ is the inverse of the Cholesky factor of $\boldsymbol{R}_0$, i.e. $\boldsymbol{\Gamma} = \boldsymbol{L}^{-1}$ where $\boldsymbol{M}$ is the unique lower triangular matrix with positive diagonal elements such that $\boldsymbol{M}\boldsymbol{M}^t = \boldsymbol{R}_0$. The numerical computation of $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ needs some precautions if $\boldsymbol{R}_0^{-1}$ is ill-conditioned. See [Hig02] for the management of the numerical stability issue.

The interest of this transformation resides in its action on a given random vector $\boldsymbol{X}$ with marginal distribution functions $F_1, \ldots, F_n$. In this case, the random vector $\boldsymbol{W} = T_1(\boldsymbol{X})$ has standard normal marginal distribution functions. At this point, without additional hypotheses, $\boldsymbol{W}$ is not in general a Gaussian vector.
In addition, if $\boldsymbol{R}_0$ is the correlation matrix of $\boldsymbol{W}$, then the random vector $\boldsymbol{U} = T_2(\boldsymbol{W}) = T(\boldsymbol{X})$ has uncorrelated components, as one can see by computing its covariance matrix:

$$\begin{aligned}
\mathbf{Cov}\left[\boldsymbol{U}\right] &= \mathbf{Cov}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{\Gamma}\boldsymbol{W} - \mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right]\right)\left(\boldsymbol{\Gamma}\boldsymbol{W} - \mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right]\right)^t\right] \\
&= \mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}(\boldsymbol{\Gamma}\boldsymbol{W})^t - \mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right](\boldsymbol{\Gamma}\boldsymbol{W})^t - \boldsymbol{\Gamma}\boldsymbol{W}\mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right]^t + \mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right]\mathbb{E}\left[\boldsymbol{\Gamma}\boldsymbol{W}\right]^t\right] \\
&= \boldsymbol{\Gamma}\mathbb{E}\left[\boldsymbol{W}\boldsymbol{W}^t\right]\boldsymbol{\Gamma}^t - \boldsymbol{\Gamma}\mathbb{E}\left[\boldsymbol{W}\right]\mathbb{E}\left[\boldsymbol{W}\right]^t\boldsymbol{\Gamma}^t \\
&= \boldsymbol{\Gamma}\mathbf{Cov}\left[\boldsymbol{W}\right]\boldsymbol{\Gamma}^t = \boldsymbol{I}_n
\end{aligned}$$

as $\mathbf{Cov}\left[\boldsymbol{W}\right] = \mathbf{Cor}\left[\boldsymbol{W}\right] = \boldsymbol{R}_0$ due to the unit variance components of $\boldsymbol{W}$ and the relation between $\boldsymbol{\Gamma}$ and $\boldsymbol{R}_0$. We deduce that $\boldsymbol{U}$ has also unit variance and uncorrelated components.

From a probabilistic modeling point of view, this transformation is used to map a random vector $\boldsymbol{X}$ with given marginal distribution functions $F_1, \ldots, F_n$ and correlation matrix $\boldsymbol{R}$ to a random vector $\boldsymbol{U}$ **that is supposed** to be a Gaussian vector in the traditional use of the Nataf transformation. This assumption encompasses two different hypotheses:

- The random vector $\boldsymbol{W}$ is a Gaussian vector with a correlation matrix $\boldsymbol{R}_0$. This correlation matrix is called the **ficticious correlation matrix** in the reliability literature, see e.g. [DM05].
- The initial correlation $\boldsymbol{R}$ can be obtained as the correlation matrix of $T_1^{-1}(\boldsymbol{W})$, which may not be possible even if $\boldsymbol{R}$ and $F_1, \ldots, F_n$ are compatible.

The first hypothesis is the most important one for the modeling, as it expresses the joint distribution of the random vector $\boldsymbol{X}$ indirectly, by specifying the joint distribution of its image $\boldsymbol{U}$ through the Nataf transformation. The formalization of this hypothesis and the study of its consequences is the subject of the remaining of this chapter.

The second hypothesis is closely linked to the Fréchet-Hoeffding Theorem 1.30. Even if $\boldsymbol{R}$ and $F_1, \ldots, F_n$ are compatible, one must check that there exists a correlation matrix $\boldsymbol{R}_0$ such that the Gaussian vector $\boldsymbol{W}$ with standard normal marginal distributions and correlation $\boldsymbol{R}_0$ is transformed into a random vector $\boldsymbol{X}$ with correlation matrix $\boldsymbol{R}$ by T^{-1}, the marginal distributions of $\boldsymbol{X}$ being equal to $F_1, \ldots, F_n$ by construction.

In general $\boldsymbol{R}_0 \neq \boldsymbol{R}$. Indeed, we have the following relation between $\boldsymbol{R}$ and $\boldsymbol{R}_0$:

$$\begin{aligned}
r_{ij} &= \mathbb{E}\left[\left(\frac{F_i^{-1}(\Phi(W_i)) - \mu_i}{\sigma_i}\right)\left(\frac{F_j^{-1}(\Phi(W_j)) - \mu_j}{\sigma_j}\right)\right] \\
&= \frac{1}{\sigma_i\sigma_j}\iint_{\mathbb{R}^2}(F_i^{-1}(\Phi(w_i)) - \mu_i)(F_j^{-1}(\Phi(w_j)) - \mu_j)\varphi_{2,r_{0ij}}(w_i, w_j)\,\mathrm{d}w_i\mathrm{d}w_j \quad (3.4)
\end{aligned}$$

where $\mu_i = \mathbb{E}[X_i]$, $\mu_j = \mathbb{E}[X_j]$, $\sigma_i = \sqrt{\mathbf{Var}[X_i]}$, $\sigma_j = \sqrt{\mathbf{Var}[X_j]}$ and $\varphi_{2,r_{0ij}}$ is the bivariate standard normal probability density function with correlation r_{0ij}:

$$\varphi_{2,r_{0ij}}(w_i, w_j) = \frac{1}{2\pi\sqrt{1-r_{0ij}^2}} \exp\left(-\frac{w_i^2 - 2r_{0ij}w_iw_j + w_j^2}{2(1-r_{0ij}^2)}\right) \tag{3.5}$$

Remark 3.3.

- We express r_{ij} as a function of r_{0ij} because at this point, the only joint distribution we know is the distribution of $\boldsymbol{W}$. We have $r_{0ij} = \mathbb{E}[W_iW_j]$ as W_i and W_j have both a standard normal distribution. Using the expression of T_1, we get $r_{0ij} = \mathbb{E}\left[\Phi^{-1}(F_i(X_i))\Phi^{-1}(F_j(X_j))\right]$ but we are unable to compute this expectation as we do not know the joint distribution of (X_i, X_j) and $\mathbb{E}\left[\Phi^{-1}(F_i(X_i))\Phi^{-1}(F_j(X_j))\right]$ is not a function of r_{ij}, F_i and F_j only in the general case.
- The computation of the coefficients r_{0ij} might be difficult for two reasons. The first one is that it involves the resolution of the integral equation (3.4), which is not guaranteed to have a solution, in particular if r_{ij} is too close to 1 or -1. The second one is that even if each coefficient r_{0ij} can be computed, there is no guarantee that the resulting matrix $\boldsymbol{R}_0$ will be symmetric definite positive.

The Nataf transformation is said to map the **physical space** where $\boldsymbol{X}$ takes its values into the **standard space** where $\boldsymbol{U}$ takes its values. The interest of the standard space is that we can rewrite the expression of the probability of failure as

$$\begin{aligned} p = \mathbb{P}(Y \geq s) &= \int_{\mathcal{D}_s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{D}_s^U} \varphi_n(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \end{aligned} \tag{3.6}$$

where the limit state function g has been transformed by T into $G = g \circ T^{-1}$ and the failure domain $\mathcal{D}_s$ into $\mathcal{D}_s^u = \{\boldsymbol{u} \in \mathbb{R}^n / G(\boldsymbol{u}) \geq s\}$, where φ_n is the probability density function of the standard n-dimensional normal distribution:

$$\varphi_n(\boldsymbol{u}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\|\boldsymbol{u}\|^2\right) \tag{3.7}$$

The first expression involves the integral of the unknown function f over a complex domain $\mathcal{D}_s$, whereas the second expression involves the integral of the known function φ_n over the complex domain $\mathcal{D}_s^u$.

The main interest of the Nataf transformation is that φ_n is a rapidly decreasing function of $\|\boldsymbol{u}\|$, which leads us to suppose that most of the contribution of $\varphi_n(\boldsymbol{u})$ to the integral (3.6) is concentrated in the vicinity of the point of $\mathcal{D}_s^u$ that is the nearest to the origin of the standard space. This point, called the *design point* and denoted P^*, is located on the hypersphere of minimal radius that is tangent to the boundary of the failure domain. It enables us to make a geometrical simplification of the failure domain $\mathcal{D}_s^u$, by modifying its boundary. The so-called FORM method is obtained by a linearization of this boundary at the design point.

We postpone the detailed presentation of the FORM method and the other various extensions such as the SORM method to the chapter 4, as we are mainly focused in this chapter on the reinterpretation of the Nataf transformation as a tool for modeling stochastic dependence.

3.2 New interpretation of the Nataf transformation through the copula theory

The Nataf transformation is the composition of two transformations T_1 and T_2, with an additional hypothesis that upon the action of T_1, the initial random vector $\boldsymbol{X}$ is mapped into a Gaussian vector $\boldsymbol{W} = T_1(\boldsymbol{X})$ with zero mean, unit marginal covariance and correlation matrix $\boldsymbol{R}_0$, i.e. with distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_0)$, then the random vector $\boldsymbol{U}$ has a standard n-dimensional normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$.

Formalizing the hypothesis underlying the Nataf transformation leads to:

Proposition 3.4. *Let $\boldsymbol{X}$ be a random vector with unknown copula $C_{\boldsymbol{X}}$, known marginal distribution functions $F_1, \dots, F_n$ and known linear correlation matrix $\boldsymbol{R}$. Assuming that this vector is mapped into a Gaussian vector $\boldsymbol{W} = T_1(\boldsymbol{X})$ with distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{r_0})$ upon the action of T_1 as defined in (3.1) is equivalent to the assumption that $C_{\boldsymbol{X}}$ is the normal copula $C^{\mathcal{N}}_{\boldsymbol{R}_0}$ parameterized by the correlation matrix $\boldsymbol{R}_0$.*

Proof. The demonstration is a direct application of the invariance of the copula by strictly increasing transformation of the components of a random vector. By definition of the normal copula as a copula of a Gaussian vector, the copula $C_{\boldsymbol{W}}$ of $\boldsymbol{W}$ is exactly the normal copula $C^{\mathcal{N}}_{\boldsymbol{R}_0}$ parameterized by $\mathbf{Cor}\,[\boldsymbol{W}] = \boldsymbol{R}_0$. Then, the transformation T_1 is bijective and its inverse is

$$T_1^{-1} : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \boldsymbol{w} & \mapsto & \boldsymbol{x} = \begin{pmatrix} F_1^{-1} \circ \Phi(w_1) \\ \vdots \\ F_n^{-1} \circ \Phi_n(w_n) \end{pmatrix} \end{array} \right. \tag{3.8}$$

This transformation only acts on the marginal distributions of $\boldsymbol{W}$, and is a strictly increasing transformation which preserves the copula of the transformed random vector (see Proposition 1.18). We conclude that $C_{\boldsymbol{X}} = C_{\boldsymbol{W}} = C^{\mathcal{N}}_{\boldsymbol{R}_0}$. □

From Definition (3.4) of the correlation matrix $\boldsymbol{R}_0$ and the expression of a bi-dimensional marginal probability density function using (1.6) and (1.12) with $k = 2$, we have

$$r_{ij} = \frac{1}{\sigma_i \sigma_j} \iint_{\mathbb{R}^2} (x_i - \mu_i)(x_j - \mu_j) c_{ij}(F_i(x_i), F_j(x_j)) f_i(x_i) f_j(x_j)\, \mathrm{d}x_i \mathrm{d}x_j \tag{3.9}$$

where c_{ij} is the probability density function of the bi-dimensional normal copula with correlation matrix $\begin{bmatrix} 1 & r_{0ij} \\ r_{0ij} & 1 \end{bmatrix}$. The fact that $C_{\boldsymbol{X}} = C^{\mathcal{N}}_{\boldsymbol{R}_0}$ shows that there is no hope to express $\boldsymbol{R}_0$ as a function of $\boldsymbol{R}$, $F_1, \dots, F_n$ and possibly other functions independent of $\boldsymbol{R}_0$, as $\boldsymbol{R}_0$ will inevitably depend on the joint distribution function of $\boldsymbol{X}$.

From a dependence modeling point of view, the use of the Nataf transformation and the claim that $\boldsymbol{W}$ or $\boldsymbol{U}$ are Gaussian vectors is equivalent to the choice of a normal copula for the joint distribution of the input random vector $\boldsymbol{X}$. This copula is parameterized by a correlation matrix $\boldsymbol{R}_0$ in such a way that the joint distribution has the given linear correlation matrix $\boldsymbol{R}$. The relation (3.9) allows us to compute $\boldsymbol{R}_0$ from $\boldsymbol{R}$ if this last matrix is compatible with both the choice of marginal distributions and the choice of a normal copula.

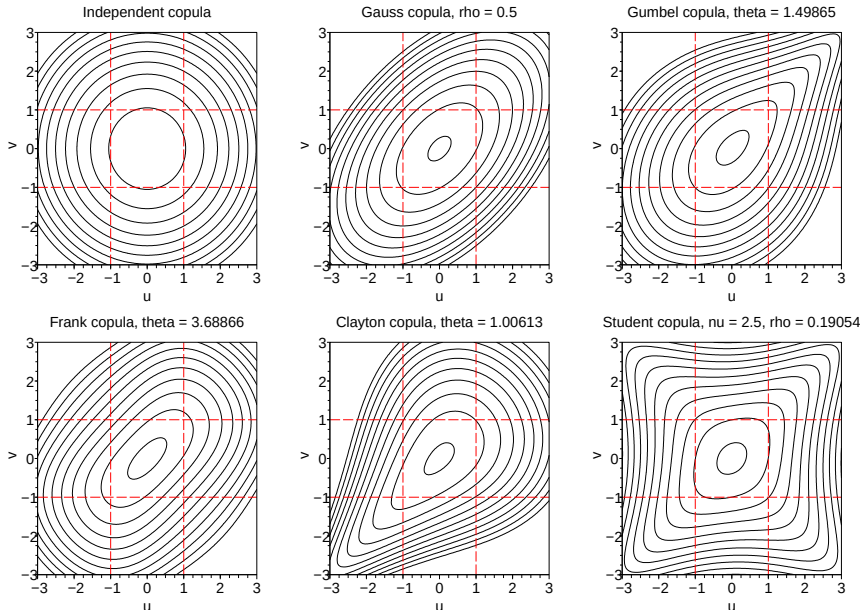


Figure 3.1: Iso-density contours of bivariate distributions with standard normal marginal distributions and different copulas such that the resulting linear correlation coefficient is equal to 0.5 (all the graphs excepted the upper-left one). We also draw the iso-contours in the case of the independent copula for comparison purpose.

3.3 Potential pitfalls of using the Nataf transformation due to the normal copula hypothesis

Thus, the use of the Nataf transformation is an obfuscated way of choosing a normal copula as a dependence structure for the input random vector. We have already seen the danger of choosing badly adapted copulas in chapter 2. Here, we present new experiments more specific to the normal copula case.

Figure 3.1 shows that with the usual available information, namely the marginal distributions and the linear correlation matrix, it is possible to choose different copulas that lead to joint distributions with exactly these characteristics, despite the visible difference between the corresponding joint probability density functions. Thus, the choice of the normal copula implies a very specific form of dependence structure, which might not suit the problem considered.

The symmetry that is visible on these graphs is only due to the specific choice of copulas taken for this illustration. A very common example of multi-dimensional distribution that arises in an industrial context such as the control of production is the case of **mixtures** of normal distributions. We give an example of such a distribution:

$$F(x_1, x_2) = \frac{1}{2}\Phi_2^1(x_1, x_2) + \frac{1}{2}\Phi_2^2(x_1, x_2) \tag{3.10}$$

where Φ_2^1 is the distribution function of the bi-dimensional normal distribution with mean vector $\boldsymbol{\mu_1} = (-0.2, 0)$, marginal standard deviations $\boldsymbol{\sigma_1} = (0.7, 0.3)$ and correlation matrix $\boldsymbol{R}_1 = \begin{bmatrix} 1 & 0.936 \\ 0.936 & 1 \end{bmatrix}$ and Φ_2^2 is the distribution function of the bi-dimensional normal distribution with mean vector $\boldsymbol{\mu_2} = (0.0, 0.2)$, marginal standard deviations $\boldsymbol{\sigma_2} = (0.8, 0.2)$ and correlation matrix $\boldsymbol{R}_2 = \boldsymbol{R}_1$.

The copula of F is obtained thanks to the Sklar theorem:

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)). \tag{3.11}$$

Then, we build the distribution $\mathcal{G}$ with copula C and standard normal marginal distributions. The distribution function G of this distribution is obtained thanks to the Sklar theorem:

$$G(x_1, x_2) = C(\Phi_{0,1}(x_1), \Phi_{0,1}(x_2)) \tag{3.12}$$

The value $\rho = 0.936$ has been chosen in such a way that G has a linear correlation of 0.8. We can see in Figure 3.2 the iso-density contours of C and G, which are clearly very different from the figure we get with a normal copula.

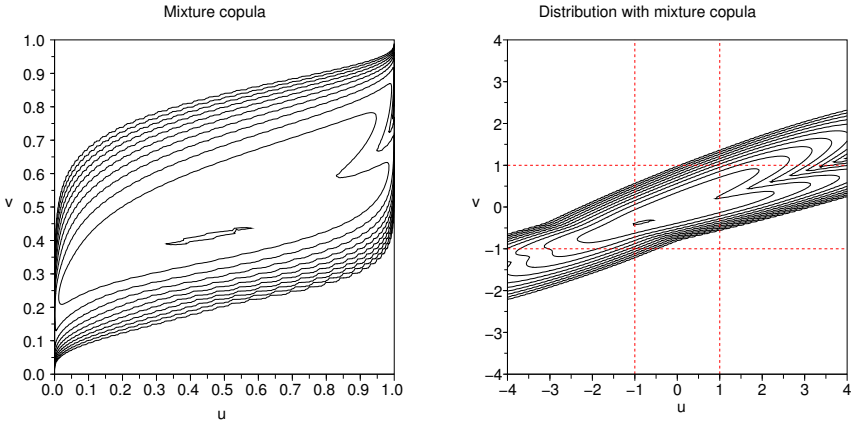


Figure 3.2: Iso-density contours of a copula of a normal mixture (left picture) and a distribution (right picture) with this copula and standard normal marginal distributions. This copula is well suited in modeling dependent quantities such as those encountered in production control.

Up to now, we have only illustrated the global effect of adopting the normal copula instead of another one on the whole support of the joint distribution. We may wonder whether these differences only affect the central part of the distribution or also modify its behaviour in the tails, which is the region we are interested in when computing low levels of probability when using first or second order reliability methods. To study this effect, we know from the experiments of Chapter 2 that the coefficient of upper tail dependence is a key parameter in the evaluation of such probabilities. From Chapter 1, for the normal copula we have $\lambda_U = \lambda_L = 0$ (Equation (1.23)), thus it is not possible to take into account any positive tail dependence with this copula, while for another copula such as the Student copula (see Table 1.1) for example, it is possible to take such tail dependence into account, as we have for this copula (Equation (1.24)):

$$\lambda_U = \lambda_L = 2 - 2T_{\nu+1}^{-1}\left(\sqrt{(1+\nu)\frac{1-\rho}{1+\rho}}\right) \tag{3.13}$$

This copula should be better suited to model the dependence structure when one is interested, for example, in a failure domain that corresponds to simultaneous large values of two components of the input random vector. To illustrate this point, we consider two dependence modelings based first on a normal copula and second on a Student copula.

These copulas share the same Spearman rho of $\rho_S = 0.01$, which leads to the linear correlation coefficients $\rho_{gauss} = 0.01047$ and $\rho_{Student} = 0.01095$ respectively for the normal copula and the Student copula. We see that the two linear correlation coefficients are very close and would be difficult to distinguish if they were estimated from real data.

The evolution of $\mathbb{P}\left(Y \geq G^{-1}(q) \mid X \geq F^{-1}(q)\right)$ with q can be seen on Figure 3.3.

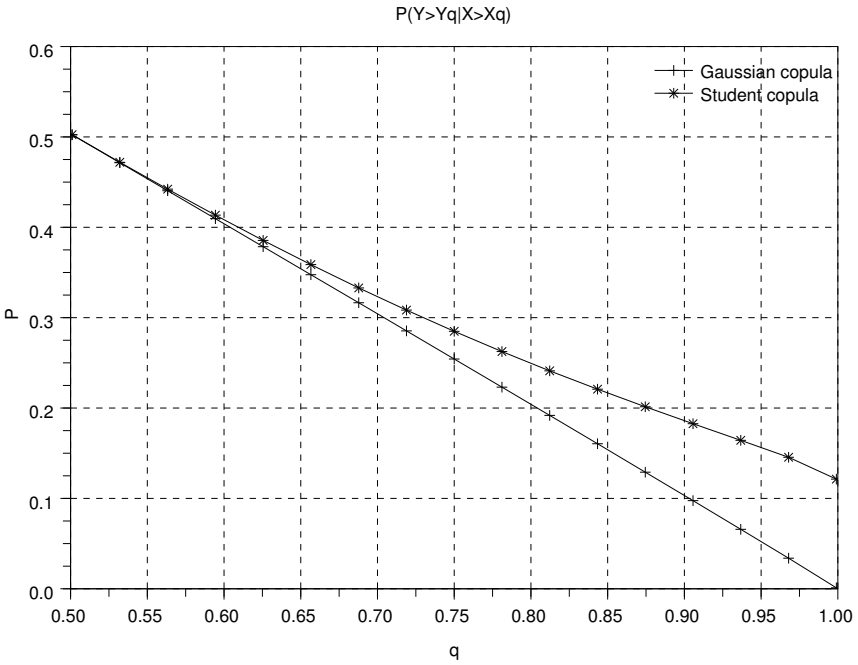


Figure 3.3: Evolution of $\mathbb{P}\left(Y \geq G^{-1}(q) \mid X \geq F^{-1}(q)\right)$ with q when X and Y are linked by the normal copula or the Student copula, with $\rho_S = 0.01$. One can see that even for moderately correlated variables, the behaviour might dramatically change in the extreme values according to the value of the tail dependence.

We can see that even for moderately correlated variables, the presence of tail dependence might dramatically change the behaviour of the distribution in its extreme values. This situation is a very typical one: when we are interested in the evaluation of a probability of failure for a system with dependent random parameters, most of the time several such parameters are in their extreme quantiles in the failure domain. Let us consider a bi-dimensional failure domain of the form

$$\mathcal{D} = \{(x, y) \mid x \geq x_q, y \geq y_q\} \tag{3.14}$$

where $x_q = F_X^{-1}(q)$, $y_q = F_Y^{-1}(q)$ and $q \simeq 1$.

If the vector (X, Y) has an upper tail dependence $\lambda_U > 0$ (see (1.22)), the probability of failure is of order $P \simeq (1-q)\lambda_U$, but otherwise it is of order $P \simeq (1-q)\varepsilon(1-q)$, where $\varepsilon(1-q) \to 0$ when $q \to 1$. This means that we can be wrong not only by a constant factor, but by several orders of magnitude if we do not take the upper tail dependence into account in our probabilistic model.

The same figure shows that the impact is much less important when we are interested in the central behaviour of the system, which means when the parameters are around their median value ($q = 0.5$ in the figure).

From these observations, it seems that even if a measure of association is by no way a full representation of the dependence structure, some such measures are better suited

to summarize this structure in the case of extremal events or in the central part of the joint distribution. This choice of measure of association has to be made in relation with the choice of a specific family of copulas. Without going too far into the problem of the selection of a copula, it is clear that the a priori restriction to a copula with no tail dependence such as the normal copula might lead to estimation of probability of failure well below their actual value if the actual copula present a positive tail dependence, with all the consequences associated to this kind of error.

3.4 Potential pitfalls of using the linear correlation to parameterize the Nataf transformation.

The tradition has consecrated the use of the linear correlation matrix as a first attempt to describe the presence of stochastic dependence. We have already seen that this choice is not optimal from the viewpoint of the notion of measure of association. We have also mentioned the two difficulties associated to the determination of the normal copula parameters from a given set of marginal distributions and a linear correlation matrix. In this section, we give more theoretical insight on these remarks. The main result upon which we will build our analysis is the Frechet-Hoeffding Theorem 1.30.

In Example 1.15, we note that ρ_{min} and ρ_{max} tend to 0 as σ goes to $+\infty$: the linear correlation between X and Y can be made as small as desired, even if Y is a strictly increasing (or decreasing) function of X (in which cases we could have expected correlations close to -1 and 1). If we restrict ourselves to the normal copulas by using the Nataf transformation, this can only emphasize this restriction. The direct consequence of this incompatibility is the impossibility to solve the equation (3.4) for some pair of components (X_i, X_j).

In industrial practices, it is common to have different teams working on the marginal distribution estimation (experts from specific physics) and the dependence modeling (experts on system modeling and interactions between systems). The modeling of the dependence resorts to the determination of a linear correlation matrix, based either on multi-dimensional data or expert judgement, with a weak link with the estimation of the marginal distribution functions. Even if the resulting marginal distribution functions and the marginal distribution functions are compatible, which is by no way enforced in the estimation process, there is no guarantee that this correlation matrix is also compatible with a normal copula.

Finally, the linear correlation matrix evaluated by experts must be positive and symmetric, with all its diagonal elements equal to 1 and the others in $[-1, 1]$. If the first properties are verified by construction of the matrix, the last one, in return, is generally not verified when the linear correlation matrix is obtained from experts. This problem becomes increasingly severe when the dimension of $\boldsymbol{X}$ grows: the set of positive matrices becomes negligible in the set of symmetric matrices with unit diagonal and off-diagonal coefficients in $[-1, 1]$ when this dimension increases.

Some of these problems can be solved by using another measure of association to parameter the dependence structure. The Spearman rho (1.16) as well as the Kendall tau (1.17) can be used. As they are functions of the copula only, they do not have to fulfill a compatibility condition with the marginal distributions. For a general copula that depends on a vector of parameters $\boldsymbol{\theta}$, one can use relations between enough measures of association and $\boldsymbol{\theta}$ to get the value of the parameters, e.g. relations (1.18) in the case of a normal copula. Using these relations, it is very easy to compute the whole correlation matrix of a multi-dimensional normal copula from a matrix of corresponding Spearman

rho or Kendall tau. Nevertheless, one must be aware of the fact that all the problems have not been fixed: one has to check that the resulting correlation matrix is positive definite. If this is not the case, it means that the given Spearman rho matrix or Kendall tau matrix is not compatible with the normal copula hypothesis.

3.5 Conclusion

The objective of this chapter was to take benefit from the copula theory to give more insight into the Nataf transformation than the presentations given so far. The central role played by this transformation in probabilistic safety assessment studies plainly justifies this need of insight.

This innovating viewpoint has enabled us to demonstrate that the Nataf transformation is a particular modeling of the stochastic dependence, using the normal copula. Furthermore, the traditional use of the Nataf transformation requires the linear correlation matrix of the input random vector in order to parameterize the normal copula.

We have shown the consequences of such an hypothesis and choice of parameters, which has enabled us to guard against the pitfalls of a systematic use of the Nataf transformation, as presented in the literature. In particular, we have showed the impact of the choice of a normal dependence structure on the morphology of the probabilistic distribution of the input random vector and on its tail dependence properties.

Furthermore, we have made explicit why using the linear correlation matrix in order to parameterize a normal copula might cause great difficulties, mainly because of the Frechet-Hoeffding theorem which constrains the linear coefficients within a range of variation depending on the marginal distributions of the random vector. In particular, this viewpoint has enabled us to understand why the application of the Nataf transformation sometimes appears impossible, which has never been explained so far.

Finally, we raised the difficulties inherent to the determination of the linear correlation matrix by expert judgements, often realized independently of the determination of the marginal distributions of the random vector.

In order to deal with these difficulties, we proposed the parameterization of the normal copula from the Spearman rho correlation matrix or the Kendal tau matrix: these measures of association are more adapted to give information on the dependence structure than the linear correlation coefficient.

Thanks to this innovating viewpoint, the Nataf transformation can be extended to more general dependence structures, namely the elliptical copulas of which the normal copula is a special case, as it will be presented in the next chapter.

Chapter 4

The generalized Nataf transformation

We saw in the previous chapter that the choice of an isoprobabilistic transformation is one of the most important steps in the use of the First Order Reliability Method (FORM), when one needs to compute an approximation of the probability of a rare event. This isoprobabilistic transformation T is a diffeomorphism from $\mathrm{supp}(\boldsymbol{X})$ into $\mathbb{R}^n$, such that the distribution of the random vector $\boldsymbol{U} = T(\boldsymbol{X})$ has the following properties: $\boldsymbol{U}$ and $\boldsymbol{Q}\boldsymbol{U}$ have the same distribution for all orthogonal transformation $\boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})$. In the previous chapter, we detailed such a transformation when $\boldsymbol{X}$ has a normal copula, namely the Nataf transformation.

The objective of this chapter is threefold: to give a quick introduction to elliptical distributions and copulas, to propose a generalization of the Nataf transformation to any random vector $\boldsymbol{X}$ whose copula is elliptical and not necessarily normal, and to provide an extension of the FORM and SORM approximations to this generalized Nataf transformation. It is a detailed exposition of the results originally presented in [Leb04] and published in an extended form in [LD09b].

4.1 Spherical and elliptical distributions

The objective of this section is to give a quick introduction to spherical and elliptical distributions. It is a key step for the presentation of the elliptical copula and the generalized Nataf transformation. The reader will find a more detailed presentation in [KFN87], as well as the proofs of the results we present. The notion of elliptical distributions can be viewed as an extension of the notion of multivariate normal distribution (which is a particular family of elliptical distributions) to families of multi-dimensional distributions that share the property of being invariant by affine transformation, the same way a multi-dimensional normal distribution remains a multivariate normal distribution after an affine transformation.

4.1.1 Spherical distributions

We first define spherical distribution, which is a step towards the definition of elliptical distribution. Three equivalent definitions are possible, depending on the viewpoint we choose. The first definition is based on the particular form of the characteristic function of such distributions. The second definition is based on invariance with respect to the group of orthogonal transformations $\mathcal{O}_n(\mathbb{R})$. The third definition relies on the stochastic

representation of any random vector whose distribution is spherical. As a preambule of these definitions, we recall some basic facts about groups of transformations, taken from [KFN87], in order to get the notion of **maximal invariant**.

We start by the notion of **group of transformations**:

Definition 4.1. Let $\mathcal{G}$ be a non-empty set of transformations from a space $\mathcal{H}$ into itself. $\mathcal{G}$ is a **group of transformations** if and only if it satisfies the following conditions:

1. If $g_1 \in \mathcal{G}$ and $g_2 \in \mathcal{G}$, then $g_1 g_2 \in \mathcal{G}$ where $g_1 g_2$ is defined as $(g_1 g_2)(x) = g_1(g_2(x))$ for all $x \in \mathcal{H}$.
2. If $g \in \mathcal{G}$, there exists a $g^{-1} \in \mathcal{G}$ such that $gg^{-1} = g^{-1}g = e$, where e is the identity transformation in $\mathcal{G}$.

Note that necessarily, $e \in \mathcal{G}$ and the inverse g^{-1} of $g \in \mathcal{G}$ is unique.

We define now the **equivalence with respect to a group of transformations**:

Definition 4.2. Two points x_1 and x_2 in $\mathcal{H}$ are said to be **equivalent** under $\mathcal{G}$ if there exists a $g \in \mathcal{G}$ such that $x_2 = gx_1$. We write $x_1 \sim x_2 \pmod{\mathcal{G}}$. This relation has the following properties:

1. $x \sim x \pmod{\mathcal{G}}$;
2. $x \sim y \pmod{\mathcal{G}}$ implies $y \sim x \pmod{\mathcal{G}}$;
3. $x \sim y \pmod{\mathcal{G}}$ and $y \sim z \pmod{\mathcal{G}}$ implies $x \sim z \pmod{\mathcal{G}}$.

The notion of **invariant** is given in the following definition:

Definition 4.3. A function f defined on $\mathcal{H}$ is said to be **invariant** under $\mathcal{G}$ if

$$\forall x \in \mathcal{H}, \forall g \in \mathcal{G}, f(g(x)) = f(x)$$

Then, we get the notion of **maximal invariant**:

Definition 4.4. A function f defined on $\mathcal{H}$ is said to be a **maximal invariant** under $\mathcal{G}$ if it is invariant under $\mathcal{G}$ and if for $x_1, x_2 \in \mathcal{H}$, $f(x_1) = f(x_2)$ implies that x_1 and x_2 are equivalent.

Considering the orthogonal group $\mathcal{O}_n(\mathbb{R})$ acting on $\mathbb{R}^n$, i.e n-dimensional square matrices $\boldsymbol{Q}$ such that $\boldsymbol{Q}^t = \boldsymbol{Q}^{-1}$, we verify that the function f defined by $f(\boldsymbol{x}) = \boldsymbol{x}^t\boldsymbol{x}$ is a maximal invariant under $\mathcal{O}_n(\mathbb{R})$. f is clearly invariant under $\mathcal{O}_n(\mathbb{R})$ as for all $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})$, $f(\boldsymbol{Qx}) = \boldsymbol{x}^t\boldsymbol{Q}^t\boldsymbol{Qx} = \boldsymbol{x}^t\boldsymbol{x} = \|\boldsymbol{x}\|^2$. Furthermore, if $\|\boldsymbol{x}_1\|^2 = f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2) = \|\boldsymbol{x}_2\|^2$, by the Grahm-Schmidt process there exists a $\boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})$ such that $\boldsymbol{x}_2 = \boldsymbol{Qx}_1$, i.e $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are equivalent; thus f is a maximal invariant under $\mathcal{O}_n(\mathbb{R})$.

We have the following theorem relating invariants and maximal invariants [KFN87, Theorem 1.1]:

Theorem 4.5. *Assume that the function f defined on $\mathcal{H}$ is a maximal invariant under $\mathcal{G}$. Then a function h defined on $\mathcal{H}$ is invariant under $\mathcal{G}$ if and only if there exists a function v from $f(\mathcal{H})$ into $h(\mathcal{H})$ such that:*

$$\forall x \in \mathcal{H}, h(x) = v(f(x)) \tag{4.1}$$

The spherical distributions are defined in terms of invariance under the orthogonal group:

Definition 4.6. A n-dimensional random vector $\boldsymbol{X}$ has a **spherical distribution** if and only if :

$$\forall \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R}), \quad \boldsymbol{X} \overset{\mathcal{D}}{=} \boldsymbol{Q}\,\boldsymbol{X} \tag{4.2}$$

i.e $\boldsymbol{X}$ and $\boldsymbol{QX}$ have the same distribution.

This definition shows that spherical distributions are also **exchangeable**. Exchangeable distributions are distributions invariant under the group $\mathfrak{S}_n(\mathbb{R})$ of n-dimensional permutations. As $\mathfrak{S}_n(\mathbb{R}) \subset \mathcal{O}_n(\mathbb{R})$, it shows that a spherical distribution is also exchangeable. The exchangeable distributions play a key role in statistics, as the joint distribution of a statistical sample must be exchangeable as soon as it does not depend on the order in which the observations are made.

When $\boldsymbol{X}$ has a spherical distribution and has also an absolutely continuous distribution function, its density is also invariant under $\mathcal{O}_n(\mathbb{R})$:

Proposition 4.7. *If the distribution function of $\boldsymbol{X}$ is absolutely continuous and its density function is $p_{\boldsymbol{X}}$, then there exists a function θ of a scalar variable such that $\forall \boldsymbol{x} \in \mathbb{R}^n$, $p_{\boldsymbol{X}}(\boldsymbol{x}) = \theta(\|\boldsymbol{x}\|^2)$. The function θ is called the* **density generator** *of the distribution. The density function $p_{\boldsymbol{X}}$ is then invariant under $\mathcal{O}_n(\mathbb{R})$.*

Proof. If $\boldsymbol{X}$ has a spherical distribution, then for all $\boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})$, $\boldsymbol{W} = \boldsymbol{QX}$ has the same distribution as $\boldsymbol{X}$ thus the same density function. Since the density functions $p_{\boldsymbol{W}}$ of $\boldsymbol{W}$ is also related to the density function $p_{\boldsymbol{X}}$ by:

$$\forall \boldsymbol{w} \in \mathbb{R}^n,\ p_{\boldsymbol{W}}(\boldsymbol{w}) = p_{\boldsymbol{X}}\left(\boldsymbol{Q}^t\boldsymbol{w}\right) \tag{4.3}$$

we have:

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \forall \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R}),\ p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{\boldsymbol{X}}\left(\boldsymbol{Q}^t\boldsymbol{x}\right) \tag{4.4}$$

which shows that $p_{\boldsymbol{X}}$ is invariant under $\mathcal{O}_n(\mathbb{R})$ and by Theorem 4.5 is a function of the maximal invariant $\|\boldsymbol{x}\|^2$. □

The definition of spherical distributions imposes also a specific form for their characteristic functions, as the following theorem (see [KFN87, Theorem 2.1]) shows:

Theorem 4.8. *A n-dimensional random vector $\boldsymbol{X}$ has a spherical distribution if and only if its characteristic function $\varphi_{\boldsymbol{X}}$ satisfies one of the following equivalent properties:*

1. $\forall \boldsymbol{t} \in \mathbb{R}^n$, $\forall \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})$, $\varphi_{\boldsymbol{X}}(\boldsymbol{Qt}) = \varphi_{\boldsymbol{X}}(\boldsymbol{t})$
2. *There exists a function ψ of a scalar variable such that $\forall \boldsymbol{t} \in \mathbb{R}^n$, $\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \psi(\|\boldsymbol{t}\|^2)$. The function ψ is called the* **characteristic generator** *of the distribution.*

The last property shows that the characteristic function is invariant under $\mathcal{O}_n(\mathbb{R})$. We note $\mathcal{S}_\psi$ the spherical distribution of $\boldsymbol{X}$ when it is characterized using its characteristic function.

Proof. For any square matrix $\boldsymbol{A}$, the characteristic function of $\boldsymbol{AX}$ equals $\varphi_{\boldsymbol{X}}(\boldsymbol{A}^t\boldsymbol{X})$, that is:

$$\mathbb{E}\left[e^{i\boldsymbol{t}^t\boldsymbol{AX}}\right] = \mathbb{E}\left[e^{i\boldsymbol{A}^t\boldsymbol{t}^t\boldsymbol{X}}\right] = \varphi_{\boldsymbol{X}}\left(\boldsymbol{A}^t\boldsymbol{t}\right) \tag{4.5}$$

thus the property 1 is equivalent to the relation 4.2. Now, the property 2 implies the property 1, since:

$$\varphi_{\boldsymbol{X}}\left(\boldsymbol{Q}^t\boldsymbol{t}\right) = \psi\left(\boldsymbol{Q}^t\boldsymbol{t}^t\boldsymbol{Q}^t\boldsymbol{t}\right) = \psi\left(\boldsymbol{t}^t\boldsymbol{Q}\boldsymbol{Q}^t\boldsymbol{t}\right) = \psi\left(\|\boldsymbol{t}\|^2\right) = \varphi_{\boldsymbol{X}}\left(\boldsymbol{t}\right) \tag{4.6}$$

Conversely, the property 1 implies that $\varphi_{\boldsymbol{X}}$ is an invariant function with respect to $\mathcal{O}_n\left(\mathbb{R}\right)$ which has the maximal invariant $\|\boldsymbol{t}\|^2$ □

This theorem allows one to prove easily the following classical result : "the only spherical distributions with independent components are the normal distributions with zero mean and covariance matrix proportional to the identity". For a demonstration, see e.g. [AL82].

The function ψ characterizes the **family** of the spherical distribution (e.g. Gaussian, Student etc.), up to a scaling factor: for any constant $c > 0$, $\boldsymbol{X}$ and $c\boldsymbol{X}$ are in the same family, which means that ψ and $\psi\left(c^2\,.\right)$ define the same family of spherical distributions.

Random vectors with spherical distributions are characterized by a specific stochastic representation, as the next theorem shows:

Definition 4.9. A n-dimensional random vector $\boldsymbol{X}$ has a spherical distribution if and only if there exists a random variate $R \geq 0$ and a random vector $\boldsymbol{U}$ independent of R and uniformly distributed on the hypersphere $\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| = 1\}$, such that:

$$\boldsymbol{X} = R\boldsymbol{U} \tag{4.7}$$

Proof. See [KFN87, Theorem 2.5] □

This representation provides also an efficient way to sample the underlying distribution, by sampling independently its radial part R and its standard spherical part $\boldsymbol{U}$. It is also very useful in order to reduce the computation of the probability content of a half-space to an univariate integration of the radial part R (which is the basis of the FORM method, see section 4.4).

The mean and the covariance of a spherical distribution exist if and only if they exist for the distribution of the associated R (see equation (4.7)). Given that $\mathbb{E}\left[\boldsymbol{U}\right] = \boldsymbol{0}$ and the independence between R and $\boldsymbol{U}$, we have :

If $\mathbb{E}\left[R\right] < \infty$,

$$\mathbb{E}\left[\boldsymbol{X}\right] = \mathbb{E}\left[R\boldsymbol{U}\right] = \mathbb{E}\left[R\right]\mathbb{E}\left[\boldsymbol{U}\right] = \boldsymbol{0} \tag{4.8}$$

If $\mathbb{E}\left[R^2\right] < \infty$,

$$\begin{aligned} \mathbf{Cov}\left[\boldsymbol{X}\right] &= \mathbb{E}\left[\boldsymbol{X}\,\boldsymbol{X}^t\right] - \mathbb{E}\left[\boldsymbol{X}\right]\mathbb{E}\left[\boldsymbol{X}\right]^t \\ &= \mathbb{E}\left[R^2\right]\mathbb{E}\left[\boldsymbol{U}\,\boldsymbol{U}^t\right] \\ &= \mathbb{E}\left[R^2\right]\mathbf{Cov}\left[\boldsymbol{U}\right] \end{aligned} \tag{4.9}$$

If we take the standard normal distribution $\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_n\right)$ for $\boldsymbol{X}$, the quantity $\|\boldsymbol{X}\|^2 = R^2$ is $\chi^2\left(n\right)$-distributed, so $\mathbb{E}\left[R^2\right] = n$ and (4.9) rewrites:

$$\begin{aligned} \mathbf{Cov}\left[\boldsymbol{X}\right] &= n\mathbf{Cov}\left[\boldsymbol{U}\right] = \boldsymbol{I}_n \\ \mathbf{Cov}\left[\boldsymbol{U}\right] &= \tfrac{1}{n}\boldsymbol{I}_n \end{aligned} \tag{4.10}$$

From (4.9) and (4.10) we deduce that for a general spherically distributed $\boldsymbol{X}$ such that $\mathbb{E}\left[R^2\right] < \infty$:

$$\mathbf{Cov}\left[\boldsymbol{X}\right] = \frac{1}{n}\mathbb{E}\left[R^2\right]\boldsymbol{I}_n \tag{4.11}$$

The definitions 4.8, 4.6 and 4.9 are equivalent according to [KFN87, Theorems 2.1, 2.2 and its corollary].

4.1.2 Elliptical distributions

We can now define the elliptical distributions:

Definition 4.10. A random vector $\boldsymbol{X}$ in $\mathbb{R}^n$ has an **elliptical distribution** if and only if there exists a deterministic vector $\boldsymbol{\mu} \in \mathbb{R}^n$, a n by p deterministic matrix $\boldsymbol{A}$, and a spherically distributed random vector $\boldsymbol{V} \in \mathbb{R}^p$ with $p = \operatorname{rank}(\boldsymbol{X}) \leq n$ such that:

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\,\boldsymbol{V} \tag{4.12}$$

where $\operatorname{rank}(\boldsymbol{X})$ is defined as the dimension of the smallest subspace of $\mathbb{R}^n$ in which $\boldsymbol{X}$ takes its values almost surely.

A (possibly degenerated) elliptically distributed random vector is thus the image of a (possibly lower dimensional) spherically distributed random vector by an affine transformation. Of course, the set of elliptical distributions contains the set of spherical distributions. Using the stochastic representation of $\boldsymbol{V}$, we get:

Proposition 4.11. *A random vector $\boldsymbol{X}$ in $\mathbb{R}^n$ has an elliptical distribution if and only it is possible to find a deterministic vector $\boldsymbol{\mu}$, a n by p matrix $\boldsymbol{A}$ with $p = \operatorname{rank}(\boldsymbol{X})$, a positive scalar random variate R and a random vector $\boldsymbol{U}$ independent of R and uniformly distributed on the unit hypersphere of $\mathbb{R}^p$, such that:*

$$\boldsymbol{X} = \boldsymbol{\mu} + R\,\boldsymbol{A}\,\boldsymbol{U}$$

which is the decomposition used for the generation of realizations of an elliptical distribution. In terms of characteristic function, the following result holds.

Proposition 4.12. *A random vector $\boldsymbol{X}$ in $\mathbb{R}^n$ has an elliptical distribution if and only if there exists a deterministic vector $\boldsymbol{\mu}$ such that the characteristic function of $\boldsymbol{X} - \boldsymbol{\mu}$ is a scalar function of the quadratic form $\boldsymbol{u}^t\boldsymbol{\Sigma}\,\boldsymbol{u}$:*

$$\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{u}) = \psi\left(\boldsymbol{u}^t\,\boldsymbol{\Sigma}\,\boldsymbol{u}\right)$$

with $\boldsymbol{\Sigma}$ a symmetric positive definite matrix of rank p. The matrix $\boldsymbol{\Sigma}$ is related to the $\boldsymbol{A}$ of Proposition 4.11 through the relation $\boldsymbol{\Sigma} = \boldsymbol{A}\,\boldsymbol{A}^t$.

We note $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma},\psi}$ the elliptical distribution of $\boldsymbol{X}$.

If the distribution of $\boldsymbol{X} - \boldsymbol{\mu}$ is continuous (which implies that $\boldsymbol{\Sigma}$ is invertible), its probability density function $p_{\boldsymbol{X}-\boldsymbol{\mu}}$ takes the form:

$$p_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{x}) = (\det\boldsymbol{\Sigma})^{-1/2}\,\theta\left(\boldsymbol{x}^t\,\boldsymbol{\Sigma}^{-1}\,\boldsymbol{x}\right) \tag{4.13}$$

where θ is a positive scalar function. The function θ is called **density generator** of the distribution, which is related to ψ by:

$$\forall t \in \mathbb{R},\ \psi(t) = \int_0^{+\infty} \Omega_n\left(tu^2\right)\theta(u)\,\mathrm{d}u \tag{4.14}$$

Elliptical distributions share many of the properties of the multivariate normal distribution (which is a special case of elliptical distributions), one of which is the algebra under affine transformation:

Proposition 4.13. *Let $\boldsymbol{X}$ in $\mathbb{R}^n$ be a random vector with distribution $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma},\psi}$, $\boldsymbol{A}$ a deterministic p by n matrix and $\boldsymbol{b}$ in $\mathbb{R}^p$ a deterministic vector. The distribution of $\boldsymbol{Y} = \boldsymbol{A}\,\boldsymbol{X}+\boldsymbol{b}$ is $\mathcal{E}_{\boldsymbol{\mu}',\boldsymbol{\Sigma}',\psi}$, where:*

$$\begin{aligned}\boldsymbol{\mu}' &= \boldsymbol{b} + \boldsymbol{A}\,\boldsymbol{\mu}\\ \boldsymbol{\Sigma}' &= \boldsymbol{A}\,\boldsymbol{\Sigma}\,\boldsymbol{A}^t\end{aligned} \tag{4.15}$$

The set of elliptical distributions in a given family characterized by the function ψ is invariant under affine transformation. In particular, the marginal distributions of a given elliptical distribution are in the same family.

Let $\boldsymbol{X}$ be a random vector that follows a given elliptical distribution and $\boldsymbol{\mu}$ and $\boldsymbol{V}$ be defined as in Definition 4.10. If the spherical distribution of $\boldsymbol{V}$ has finite mean and covariance, the elliptical distribution has finite mean and covariance too and we have:

$$\mathbb{E}\left[\boldsymbol{X}\right] = \mathbb{E}\left[\boldsymbol{\mu} + \boldsymbol{A}\,\boldsymbol{V}\right] = \boldsymbol{\mu} + \boldsymbol{A}\,\mathbb{E}\left[\boldsymbol{V}\right] = \boldsymbol{\mu} \tag{4.16}$$

and

$$\mathbf{Cov}\left[\boldsymbol{X}\right] = \mathbf{Cov}\left[\boldsymbol{\mu} + \boldsymbol{A}\,\boldsymbol{V}\right] = \boldsymbol{A}\,\mathbf{Cov}\left[\boldsymbol{V}\right]\boldsymbol{A}^t = \frac{1}{n}\mathbb{E}\left[R^2\right]\boldsymbol{\Sigma} \tag{4.17}$$

The probabilistic distribution of R characterizes the type of elliptical distribution. For example, for a normal distribution of dimension n, R^2 follows a χ^2 distribution with n degrees of freedom.

Remark 4.14. The expression in Proposition 4.11 shows that the pair $(R, \boldsymbol{A})$ is defined up to a multiplicative constant. If $\mathbb{E}\left[R^2\right] < \infty$, we will assume that this constant has been chosen so that $\mathbb{E}\left[R^2\right] = n$. Then, $\boldsymbol{\Sigma}$ is exactly the covariance matrix of $\boldsymbol{X}$. If $\mathbb{E}\left[R^2\right] = \infty$, one can choose the constant such that R^2 has the same median as a $\chi^2(n)$ distribution. Whatever the normalization is, the pair $(R, \boldsymbol{A})$ is uniquely defined for an elliptically distributed random vector, and the pair $(\boldsymbol{\Sigma}, \psi)$ is uniquely defined for the associated distribution. We will assume that such a normalization has been made for the remaining of this chapter.

As $\boldsymbol{\Sigma}$ is a positive semidefinite symmetric matrix, it can be written in the form $\boldsymbol{\Sigma} = \boldsymbol{D}\,\boldsymbol{R}\,\boldsymbol{D}$, where $\boldsymbol{D}$ is the diagonal matrix $\mathbf{diag}\,\boldsymbol{\sigma_i}$ with $\forall i \in \{1,\ldots,n\}$, $\sigma_i = \sqrt{\Sigma_{ii}} \geq 0$. We note $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_n)$. If the distribution has a covariance matrix, with our choice of normalization for ψ, we know that this covariance matrix is equal to $\boldsymbol{\Sigma}$. The matrix $\boldsymbol{R}$, such that $R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$, is then its linear correlation matrix. This matrix is well-defined,

even if the distribution has no finite second moments. In all the cases, we call it the correlation matrix of the distribution.

To summarize, an elliptical distribution is fully characterized by its **location parameter** $\boldsymbol{\mu}$, equal to the mean of the elliptical distribution if it has a finite first moment, its **marginal scale parameter** $\boldsymbol{\sigma}$, equal to the vector of standard deviations of the one-dimensional marginal distributions if they have finite second moments, its **correlation matrix** $\boldsymbol{R}$, equal to the linear correlation matrix of the elliptical distribution if it has finite second moments, and its **characteristic generator** ψ, which is a positive scalar function that characterizes the type of the elliptical distribution.

From now on, we will denote by $E_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{R},\psi}$ the distribution function of $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{R},\psi}$, corresponding to the previous notation $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma},\psi}$ where $\boldsymbol{\sigma} = \boldsymbol{D}\,\boldsymbol{R}\,\boldsymbol{D}$.

4.2 Spherical and elliptical Copulas, generic elliptical representative, standard spherical representative

Definition 4.15. An **elliptical copula** $C^E_{\boldsymbol{R},\psi}$ is the copula of an elliptical distribution $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{R},\psi}$.

Remark 4.16. Thanks to the normalization presented in Remark 4.14, the mapping $(\boldsymbol{R},\psi) \mapsto C^E_{\boldsymbol{R},\psi}$ is one-to-one. The **type** of the copula is given by ψ and its **shape** by $\boldsymbol{R}$.

In general, the copula $C^E_{\boldsymbol{R},\psi}$ is **not** the distribution function of an elliptical distribution itself.

It is clear that mapping between the elliptical distributions and elliptical copulas is not one-to-one. Let $\mathcal{R}$ be the equivalence relation between elliptical distributions: $\mathcal{E}_{\boldsymbol{\mu}_1,\boldsymbol{\sigma}_1,\boldsymbol{R}_1,\psi_1} \overset{\mathcal{R}}{\equiv} \mathcal{E}_{\boldsymbol{\mu}_2,\boldsymbol{\sigma}_2,\boldsymbol{R}_2,\psi_2}$ if and only if $\mathcal{E}_{\boldsymbol{\mu}_1,\boldsymbol{\sigma}_1,\boldsymbol{R}_1,\psi_1}$ and $\mathcal{E}_{\boldsymbol{\mu}_2,\boldsymbol{\sigma}_2,\boldsymbol{R}_2,\psi_2}$ share the same copula $C^E_{\boldsymbol{R},\psi}$. From Remark 4.16, this relation reads $(\boldsymbol{R}_1,\psi_1) = (\boldsymbol{R}_2,\psi_2)$.

We introduce the notion of generic representative to distinguish one particular elliptical distribution in each class of equivalence:

Definition 4.17. The **generic elliptical representative** of an elliptical distribution family $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{R},\psi}$ through the equivalence relation $\mathcal{R}$ is the elliptical distribution whose distribution function is $E_{\boldsymbol{0},\boldsymbol{1},\boldsymbol{R},\psi}$.

All other members of the equivalence class differ only by their location parameter $\boldsymbol{\mu}$ and their marginal scale parameter $\boldsymbol{\sigma}$.

We introduce a last kind of elliptical distributions that allows one to focus on the type of a elliptical distribution, throwing away the shape information.

Definition 4.18. The **standard spherical representative** of an elliptical distribution family $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{R},\psi}$ is $\mathcal{S}_\psi$, the spherical distribution whose distribution function is $S_\psi = E_{\boldsymbol{0},\boldsymbol{1},\boldsymbol{I}_n,\psi}$.

It is the only member of the elliptical family which is both spherical and with null location parameter and unit marginal scale parameter.

Definition 4.19. The family of distributions with marginal distribution functions $F_1,\dots,F_n$ and elliptical copula $C^E_{\boldsymbol{R},\psi}$ is denoted by $\mathcal{D}_{F_1,\dots,F_n,C^E_{\boldsymbol{R},\psi}}$. The distribution function of this distribution is denoted $D_{F_1,\dots,F_n,C^E_{\boldsymbol{R},\psi}}$.

The relationship between the different kinds of elliptical distributions is depicted in Figure 4.1. We also show how general distributions with elliptical copulas interact with these distributions.

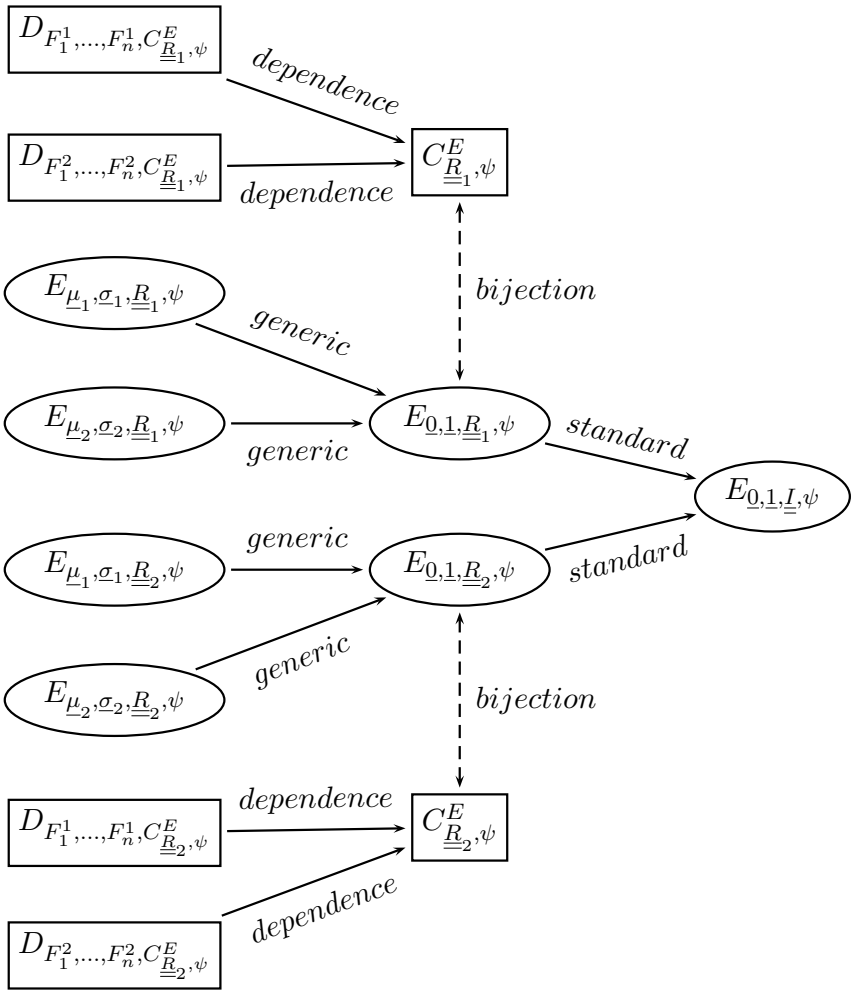


Figure 4.1: Graph showing the relations between the several kinds of elliptical and spherical distributions (oval nodes), and how general distributions with elliptical copulas (rectangular nodes) are linked to these distributions through bijection between elliptical copulas and generic elliptical representatives. The labels on the links are related to what is extracted from the left-hand side to go to the right-hand side. For example, extraction of the dependence structure of a general distribution with elliptical copula leads to its copula.

4.3 Generalized Nataf transformation

The usual Nataf transformation [Nat62] has already been analysed in the light of the copula theory in the previous chapter. It transforms a random vector into a multivariate standard Gaussian vector if and only if the copula of the random vector is normal, and some of the consequences of such an hypothesis have been presented in the context of the evaluation of low probabilities of failure. Here, we propose a generalization of this transformation to a random vector with an elliptical copula $C^E_{\boldsymbol{R},\psi}$. In this section, the random vector $\boldsymbol{X}$ is supposed to be continuous and with full rank. We also suppose that its marginal distribution functions are strictly increasing (so they are bijective) and that the matrix $\boldsymbol{R}$ of its elliptical copula is symmetric positive definite.

The usual Nataf transformation has been described as the composition of two transformations T_1 and T_2 in (3.1). The transformation T_1 can also be decomposed into two elementary transformations T_1' and T_1'' defined by:

$$T_1' : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \boldsymbol{x} & \mapsto & \boldsymbol{v} = \begin{pmatrix} F_1(x_1) \\ \vdots \\ F_n(x_n) \end{pmatrix} \end{array} \right. \tag{4.18}$$

which is the usual probabilistic integral transformation, and

$$T_1'' : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \boldsymbol{v} & \mapsto & \boldsymbol{w} = \begin{pmatrix} \Phi^{-1}(v_1) \\ \vdots \\ \Phi^{-1}(v_n) \end{pmatrix} \end{array} \right. \tag{4.19}$$

which leads to the decomposition $T = T_2 \circ T_1'' \circ T_1'$ for the usual Nataf transformation. It has been shown in Chapter 3, Proposition 3.4 that if $\boldsymbol{X}$ has a normal copula, the distribution of $\boldsymbol{U} = T(\boldsymbol{X})$ is the standard n-dimensional normal distribution, namely the standard spherical representative associated with $C^E_{\boldsymbol{R},\psi_{\mathcal{N}}}$. The transformation T_1' maps $\boldsymbol{X}$ into a random vector $\boldsymbol{V}$ whose distribution is the normal copula $C^E_{\boldsymbol{R},\psi_{\mathcal{N}}}$, the transformation T_1'' maps $\boldsymbol{V}$ into a random vector $\boldsymbol{W}$ whose distribution is the generic normal representative associated with $C^E_{\boldsymbol{R},\psi_{\mathcal{N}}}$ and T_2 maps $\boldsymbol{W}$ into a random vector whose distribution is the standard normal representative associated to $C^E_{\boldsymbol{R},\psi_{\mathcal{N}}}$. The $\boldsymbol{U}$-space is called the **standard space** whereas the $\boldsymbol{X}$-space is called the **physical space**. With this point of view, a natural generalization of the Nataf transformation is the following:

Definition 4.20. Let $\boldsymbol{X}$ in $\mathbb{R}^n$ be a continuous random vector following the distribution $D_{F_1,\ldots,F_n,C^E_{\boldsymbol{R},\psi}}$. The **generalized Nataf transformation** $T^{gen} : \mathbb{R}^n \to \mathbb{R}^n$ is defined by:

$$T^{gen} = T_2 \circ T_1''^{gen} \circ T_1' \tag{4.20}$$

where the transformations $T_1''^{gen}$ is given by:

$$T_1''^{gen} : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \boldsymbol{v} & \mapsto & \boldsymbol{w} = \begin{pmatrix} E^{-1}(v_1) \\ \vdots \\ E^{-1}(v_n) \end{pmatrix} \end{array} \right. \tag{4.21}$$

where E is the distribution function of standard one-dimensional elliptical distribution with characteristic generator ψ and $\boldsymbol{\Gamma}$ is the inverse of the Cholesky factor of $\boldsymbol{R}$ in Definition 3.1 of T_2, which was supposed to be nonsingular.

This transformation differs from the usual one by its second step, which is modified such that the distribution of $\boldsymbol{W} = T_1''^{gen} \circ T_1'(\boldsymbol{X})$ is the generic elliptical representative associated with the copula of $\boldsymbol{X}$. The step T_2 maps this distribution into its standard representative, following exactly the same algebra as the normal copula. In the special case where the distribution of $\boldsymbol{X}$ is already elliptical, with distribution function $E_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{R},\psi}$, the generalized Nataf transformation is an affine transformation: the transformation $T_1''^{gen} \circ T_1'$ maps the elliptical distribution into its generic representative, which is an affine transformation of each component, and the transformation T_2 is linear, thus $T^{gen} = T_2 \circ T_1''^{gen} \circ T_1'$ is affine. More precisely, if we note $\boldsymbol{V} = T_1''^{gen} \circ T_1'(\boldsymbol{X})$, we have $V_i = (X_i - \mu_i)/\sigma_i$ for all $i \in \{1, \ldots, n\}$. The generalized Nataf transformation can then be expressed in this case as:

$$\boldsymbol{U} = T^{gen}(\boldsymbol{X}) = \boldsymbol{S}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{4.22}$$

where $\boldsymbol{S}$ is the inverse of the Cholesky factor of $\boldsymbol{\Sigma} = \boldsymbol{D}\,\boldsymbol{R}\,\boldsymbol{D}$.

4.4 FORM and SORM approximations

Once the Nataf transformation has been extended to elliptical distributions, it is necessary to provide an extension of the FORM (First Order Reliability Method) and SORM (Second order Reliability Method) approximations to make the evaluation of the probability of exceedance possible. We start by recalling what these approximations are.

Given a numerical model $f : \mathbb{R}^n \to \mathbb{R}$ and a threshold $s \in \mathbb{R}$, the evaluation of the probability:

$$P_f = \mathbb{P}(f(\boldsymbol{X}) \geq s) = \int_{\mathbb{R}^n} \mathbf{1}_{f(\boldsymbol{x}) \geq s}\, p_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{4.23}$$

where $p_{\boldsymbol{X}}$ is the probability density function of $\boldsymbol{X}$, can be transformed into the evaluation of the probability:

$$P_f = \mathbb{P}(G(\boldsymbol{U}) \geq s) = \int_{\mathbb{R}^n} \mathbf{1}_{G(\boldsymbol{u}) \geq s}\, p_{\boldsymbol{U}}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \tag{4.24}$$

using a suitable change of variable $\boldsymbol{u} = T(\boldsymbol{x})$ such that the transformed random vector $\boldsymbol{U} = T(\boldsymbol{X})$ has a spherical distribution. Here, $G = f \circ T^{-1}$ is the **standard model** and $p_{\boldsymbol{U}}$ is the density function of $\boldsymbol{U}$. The vector $\boldsymbol{U}$ is said to take values in the **standard space**, whereas $\boldsymbol{X}$ is said to take values in the **physical space**. To be a proper transformation, T must be defined on the support $\operatorname{supp} \boldsymbol{X} = \overline{\{\boldsymbol{x} \in \mathbb{R}^n \mid p_{\boldsymbol{X}}(\boldsymbol{x}) > 0\}}$ of $\boldsymbol{X}$ and be a continuously differentiable function from the interior of $\operatorname{supp} \boldsymbol{X}$ into the interior of $T(\operatorname{supp} \boldsymbol{X})$ such that its inverse is also continuously differentiable.

One may wonder if such transformations exist. If $\boldsymbol{X}$ has a normal copula, the Nataf transformation is such a transformation, and we showed that if $\boldsymbol{X}$ has a general elliptical distribution the generalized Nataf transformation is such a transformation. For the most general case, the Rosenblatt transformation that will be presented in the next chapter is such a transformation. The interest of such a transformation is that $\boldsymbol{U}$ has a spherical distribution. If $p_{\boldsymbol{U}}$ (i.e its density generator θ) is a decreasing function of $\|\boldsymbol{u}\|$ in the **failure**

domain $\mathcal{D} = \{\boldsymbol{u} \,|\, G(\boldsymbol{u}) \geq s\}$, the integral in (4.24) can be approximated by an integral of $p_{\boldsymbol{U}}$ over an approximate domain $\widetilde{\mathcal{D}}$ close to $\mathcal{D}$. This domain is usually obtained by an approximation of the boundary $\partial\mathcal{D}$ of $\mathcal{D}$ at the vicinity of the point $\boldsymbol{u}^* \in \mathcal{D}$ with minimal norm, i.e. with maximum density (such a point is called the **design point**). A **linear** approximation of $\partial\mathcal{D}$ at $\boldsymbol{u}^*$ leads to the **FORM** approximation, whereas a **quadratic** one leads to the **SORM** approximation.

The components of the design point play also an important role in the sensitivity analysis of the failure in the standard space. In this space, as $\boldsymbol{u} \to \boldsymbol{u}^*$, the standard model G writes:

$$G(\boldsymbol{u}) = s + \boldsymbol{\nabla}^t G(\boldsymbol{u}^*)(\boldsymbol{u} - \boldsymbol{u}^*) + \frac{1}{2}(\boldsymbol{u} - \boldsymbol{u}^*)^t \boldsymbol{\nabla}^2 G(\boldsymbol{u}^*)(\boldsymbol{u} - \boldsymbol{u}^*) + o\left(||\boldsymbol{u} - \boldsymbol{u}^*||^2\right) \quad (4.25)$$

provided G is twice continuously differentiable at $\boldsymbol{u}^*$. By definition of $\boldsymbol{u}^*$:

$$\boldsymbol{u}^* = \underset{\boldsymbol{u} \in \mathcal{D}}{\operatorname{argmin}} ||\boldsymbol{u}||^2 \quad (4.26)$$

it exists $\lambda \in \mathbb{R}\backslash\{0\}$ such that $\nabla G(\boldsymbol{u}^*) = \lambda \boldsymbol{u}^*$. Using $\mathbf{Var}[U_i] = 1$, we get:

$$\mathbf{Var}[G(\boldsymbol{U})] = \lambda^2 \sum_{i=1}^{n} (u_i^*)^2 \mathbf{Var}[U_i] + o(1) = \lambda^2 \sum_{i=1}^{n} (u_i^*)^2 + o(1) \text{ as } \left\|\boldsymbol{\nabla}^2 G(\boldsymbol{u}^*)\right\| \to 0$$

The **reliability index** β is defined by:

$$\beta = ||\boldsymbol{u}^*||$$

The **importance factors** defined by:

$$\mathcal{F}_i = \frac{(u_i^*)^2}{\sum_{i=1}^{n} (u_i^*)^2} = \frac{(u_i^*)^2}{\beta^2}$$

are the relative contributions of the marginal variances of $\boldsymbol{U}$ to the variance of $G(\boldsymbol{U})$.

In order to derive the FORM and SORM approximations of the probability of failure P_f, using the invariance of the standard distribution by orthogonal transformation of $\boldsymbol{U}$, we can suppose that the design point $\boldsymbol{u}^*$ has components $(0, \ldots, 0, \beta)$, see Figure 4.2.

The generalized FORM approximation is obtained by a linear approximation of the boundary $\partial\mathcal{D}$ which has the form of the hyperplane tangent to $\partial\mathcal{D}$ of $\mathcal{D}$ at $\boldsymbol{u}^*$. In this case, the generalized FORM approximation of the probability of failure is:

$$P_{FORM}^{gen} = \mathbb{P}(U_1 \geq \beta) = 1 - E(\beta) = E(-\beta) \quad (4.27)$$

where E is the distribution function of the 1-D standard elliptical representative of the same type than the copula of $\boldsymbol{X}$. We recover the standard FORM approximation when $\boldsymbol{X}$ has a normal copula, as in this case $E \equiv \Phi$, the distribution function of the standard one-dimensional normal distribution.

For the generalized SORM approximation, more work is required. As in the case of a normal copula (the usual Nataf transformation), the expression of the probability of failure has no simple analytical formulation. The generalization of the Tvedt exact formula (see [Tve88], [Tve90]) does not seem to extend easily to the more general context we study here, as its proof relies on the independence of the components of the standard spherical representative, which occurs only in the normal case. But it is possible to generalize

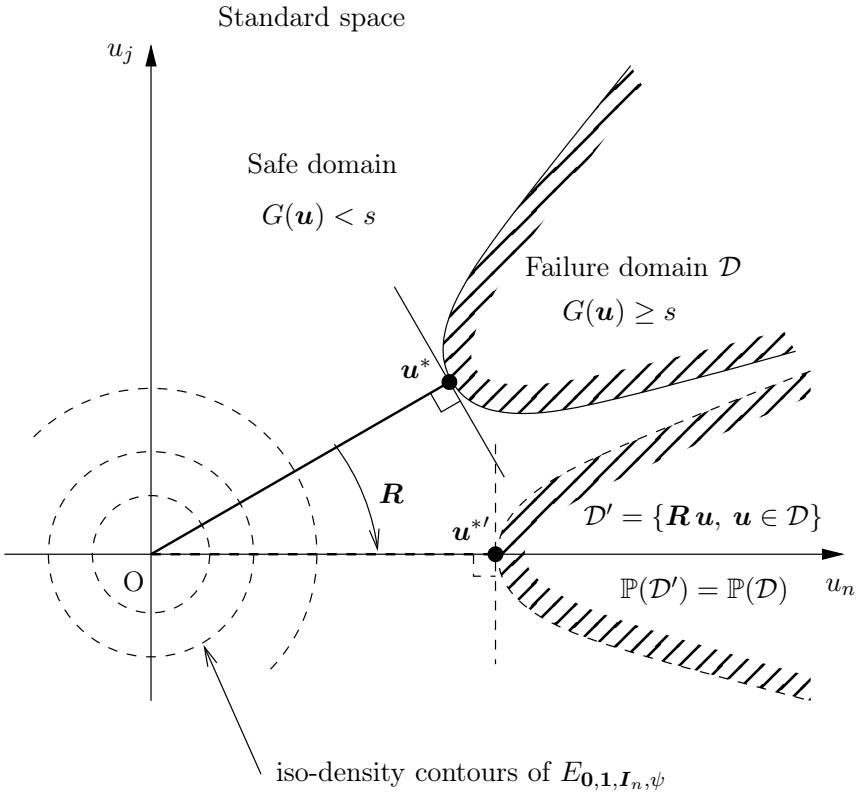


Figure 4.2: Rotational invariance after the application of the generalized Nataf transformation. The rotational invariance of $\mathcal{E}_{\mathbf{0},\mathbf{1},\boldsymbol{I}_n,\psi}$ allows one to focus on the situation depicted in dashed form without loss of generality, thanks to the rotation $\boldsymbol{R}$ that maps a general failure domain $\mathcal{D}$ to a domain $\mathcal{D}'$ for which the design point $\boldsymbol{u}^{*\prime}$ is supported by the last axis.

Breitung's asymptotic analysis derived in [Bre84] and [Bre94] to a more general spherical case. We first derive such a generalization, then we comment its traditional use that leads to Breitung's formula. We reinterpret the resulting asymptotic approximation in terms of standard models G obtained as a nonlinear transformation of a reference domain, in order to get an asymptotically exact approximation of the probability of failure of such domains.

Breitung's approximation is based on the application of the multivariate Laplace method to the evaluation of a parametric integral close to the integral in (4.24). The application of this method requires some technical conditions known as the **compactification conditions**, see [Bre94, Lemma 38]:

Lemma 4.21. *Let $\mathcal{F} \subset \mathbb{R}^n$ be a closed set and $f, h : \mathbb{R}^n \to \mathbb{R}$ be two continuous functions. Assume further that:*

1. *The set $\mathcal{M} = \{\boldsymbol{y} \in \mathcal{F} \,|\, f(\boldsymbol{y}) = \max_{\boldsymbol{x}\in\mathcal{F}} f(\boldsymbol{x})\}$ is compact.*
2. *$\int_{\mathcal{F}} |h(\boldsymbol{x})| e^{f(\boldsymbol{x})} \,\mathrm{d}\boldsymbol{x} < \infty$.*
3. *For every neighborhood $\mathcal{V}$ of $\mathcal{M}$, $\sup_{\boldsymbol{y}\in\mathcal{F}\setminus\mathcal{V}} \{f(\boldsymbol{y})\} < \max_{\boldsymbol{x}\in\mathcal{F}} f(\boldsymbol{x})$.*
4. *There exists a neighborhood $\mathcal{U}$ of $\mathcal{M}$ such that for all $\boldsymbol{x} \in \mathcal{U}$ always $h(\boldsymbol{x}) > 0$ (or $h(\boldsymbol{x}) < 0$).*
5. *For all neighborhood $\mathcal{V}$ of $\mathcal{M}$ always $\int_{\mathcal{F}\cap\mathcal{V}} \mathrm{d}\boldsymbol{x} > 0$.*

Then for all $\delta > 1$:

$$\int_{\mathcal{F}} |h(\boldsymbol{x})| e^{\delta f(\boldsymbol{x})} \,\mathrm{d}\boldsymbol{x} < \infty$$

and for all neighborhood $\mathcal{V}$ of $\mathcal{M}$, as $\delta \to \infty$:

$$\int_{\mathcal{F}} h(\boldsymbol{x}) e^{\delta f(\boldsymbol{x})} \,\mathrm{d}\boldsymbol{x} = (1 + o(1)) \int_{\mathcal{F}\cap\mathcal{V}} h(\boldsymbol{x}) e^{\delta f(\boldsymbol{x})} \,\mathrm{d}\boldsymbol{x}$$

This lemma gives conditions under which we can replace an integral over a possibly non-compact set $\mathcal{F}$ by an integral over a compact set $\mathcal{F} \cap \mathcal{V}$ without changing its asymptotic behavior, by choosing a compact neighborhood $\mathcal{V}$ of $\mathcal{M}$. It leads to the following hypothesis [Bre94, Condition A] for asymptotic approximation:

Hypothesis 4.22. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function such that $\mathcal{F} = \{\boldsymbol{x} \,\|\, g(\boldsymbol{x}) \le 0\}$ is a compact set and its boundary $\partial\mathcal{F} = \{\boldsymbol{x} \,\|\, g(\boldsymbol{x}) = 0\}$ a compact C^2 hypersurface. The gradient $\boldsymbol{\nabla} g(\boldsymbol{x})$ does not vanish on $\partial\mathcal{F}$ and the surface $\partial\mathcal{F}$ is oriented by the normal field $\boldsymbol{n}(\boldsymbol{x}) = \boldsymbol{\nabla} g(\boldsymbol{x}) / ||\boldsymbol{\nabla} g(\boldsymbol{x})||$.*

We get the following theorem, which is the root of the second order approximation of the probability of failure known as the Breitung formula, see [Bre84] and [Bre94, Theorem 46]:

Theorem 4.23. *Let Hypothesis 4.22 be fulfilled. Let $\ell : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and $h : \mathbb{R}^n \to \mathbb{R}$ be a continuous function.*

Assume that the following conditions are satisfied

1. *The function ℓ attains its global maximum with respect to $\mathcal{F}$ only at the point $\boldsymbol{x}^* \in \partial\mathcal{F}$.*
2. *The gradient of ℓ at $\boldsymbol{x}^*$ does not vanish: $\boldsymbol{\nabla}\ell(\boldsymbol{x}^*) \neq \boldsymbol{0}$.*

3. *The matrix* $\boldsymbol{H}(\boldsymbol{x}^*) = \boldsymbol{\nabla}^2\ell(\boldsymbol{x}^*) - \frac{||\boldsymbol{\nabla}\ell(\boldsymbol{x}^*)||}{||\boldsymbol{\nabla}g(\boldsymbol{x}^*)||}\boldsymbol{\nabla}^2 g(\boldsymbol{x}^*)$ *is such that the* $(n-1)\times(n-1)$ *matrix* $\boldsymbol{H}^*(\boldsymbol{x}^*) = \boldsymbol{A}(\boldsymbol{x}^*)^t \boldsymbol{H}(\boldsymbol{x}^*)\boldsymbol{A}(\boldsymbol{x}^*)$ *is nonsingular, where:*

$$\boldsymbol{\nabla}^2\phi(\boldsymbol{x}^*) = \left(\boldsymbol{\nabla}^2\phi(\boldsymbol{x}^*)\right)_{ij} = \frac{\partial^2\phi}{\partial x_i \partial x_j}(\boldsymbol{x}^*)$$

is the Hessian matrix of ϕ *at* $\boldsymbol{x}^*$*, for* $\phi \in \{g, \ell\}$*,* $\boldsymbol{A}(\boldsymbol{x}^*) = (\boldsymbol{a}_1(\boldsymbol{x}^*), \ldots, \boldsymbol{a}_{n-1}(\boldsymbol{x}^*))$ *and the vectors* $\boldsymbol{a}_1(\boldsymbol{x}^*), \ldots, \boldsymbol{a}_{n-1}(\boldsymbol{x}^*)$ *form an orthonormal basis of the tangential space of* $\mathcal{F}$ *at* $\boldsymbol{x}^*$*.*

Then, as $\delta \to \infty$ *we have the following asymptotic approximation:*

$$\int_{\mathcal{F}} h(\boldsymbol{x})\, e^{\delta^2 \ell(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} = (2\pi)^{(n-1)/2} \frac{h(\boldsymbol{x}^*)}{\sqrt{|J(\boldsymbol{x}^*)|}} \frac{e^{\delta^2 h(\boldsymbol{x}^*)}}{\delta^{n+1}} (1 + o(1)) \tag{4.28}$$

where $J(\boldsymbol{x}^*) = \boldsymbol{\nabla}\ell(\boldsymbol{x}^*)^t \boldsymbol{C}(\boldsymbol{x}^*)\boldsymbol{\nabla}\ell(\boldsymbol{x}^*)$ *and* $\boldsymbol{C}(\boldsymbol{x}^*) =$ *matrix of cofactors of* $\boldsymbol{H}(\boldsymbol{x}^*)$*.*

We can use this theorem to derive an asymptotic expansion of the probability content of a failure domain $\mathcal{D}$ which is homothetic to a reference domain $\mathcal{F}$ at unit distance from the origin, it means $\mathcal{D} = \beta\mathcal{F}$. The resulting approximation is given in the next proposition.

Proposition 4.24. *Let* G *be a standard model and* $\boldsymbol{U}$ *a random vector with a spherical distribution of density generator* θ *such that:*

1. θ *is a decreasing function;*
2. $\phi = \log(\theta)$ *is separable in the following sense:*

$$\forall \beta \in \mathbb{R}^+, \forall \boldsymbol{x} \in \mathbb{R}^n, \quad \phi\left(\beta^2 \|\boldsymbol{x}\|^2\right) = \eta(\beta)\,\ell(\boldsymbol{x}) \tag{4.29}$$

where $\eta : \mathbb{R}^+ \to \mathbb{R}$ *and* $\ell : \mathbb{R}^n \to \mathbb{R}$ *are two twice continuously differentiable functions such that:*

   (a) $\lim_{\beta\to\infty} \eta(\beta) = \infty$*;*

   (b) ℓ *satisfies the conditions of Theorem 4.23.*

Then, the asymptotic relation 4.28 reads:

$$\mathbb{P}(\beta\mathcal{F}) = \left(-\frac{\pi\theta(\beta^2)}{\theta'(\beta^2)}\right)^{(n+1)/2} \frac{\theta(\beta^2)}{2\pi\beta} \left(\prod_{i=1}^{n-1} \frac{1}{\sqrt{1-\kappa_i}}\right)(1 + o(1)) \quad \text{as } \beta \to \infty \tag{4.30}$$

where $\boldsymbol{x}^* = \mathrm{argmin}_{\boldsymbol{x}\in\mathcal{F}} ||\boldsymbol{x}||$*,* $(\kappa_i)_{i=1,\ldots,n-1}$ *are the main curvatures of* $\partial\mathcal{F}$ *at* $\boldsymbol{x}^*$ *and* θ' *is the first derivative of* θ*.*

Proof. The failure domain $\mathcal{D}$ is defined by $\mathcal{D} = \{\boldsymbol{u} \,|\, G(\boldsymbol{u}) \geq s\}$. The hypothesis that $\mathcal{D} = \beta\mathcal{F}$ with $\boldsymbol{x}^* = \mathrm{argmin}_{\boldsymbol{x}\in\mathcal{F}} \|\boldsymbol{x}\|$ and $\|x^*\| = 1$ is equivalent to the hypothesis that the standard model G is the member G_β of a parametric family of models $\mathcal{G}_\beta$ that writes:

$$\forall u \in \mathbb{R}^n, \quad G_\beta(\boldsymbol{u}) = s - g\left(\beta^{-1}\boldsymbol{u}\right) \tag{4.31}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a given fixed function and $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n \,|\, g(\boldsymbol{x} \leq 0)\}$.

We look for an asymptotic expansion when $\beta \to \infty$ of:

$$\begin{aligned}
\mathbb{P}(\beta\mathcal{F}) =& \int_{\beta\mathcal{F}} \theta\left(\|\boldsymbol{u}\|^2\right) \mathrm{d}\boldsymbol{u} \\
=& \beta^n \int_{\mathcal{F}} \theta\left(\beta^2\|\boldsymbol{x}\|^2\right) \mathrm{d}\boldsymbol{x} \\
=& \beta^n \int_{\mathcal{F}} e^{\eta(\beta)\ell(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}
\end{aligned}$$

Due to the rotational invariance of θ with respect to a rotation around the origin, we can take $\boldsymbol{x}^* = \boldsymbol{e}_n$. Then, replacing δ by $\eta(\beta)$ in Equation (4.28), we have as $\beta \to \infty$:

$$\mathbb{P}(\beta\mathcal{F}) = \beta^n (2\pi)^{(n-1)/2} \frac{e^{\eta(\beta)\ell(\boldsymbol{e}_n)}}{\eta(\beta)^{(n+1)/2}} \frac{1}{\sqrt{|J(\boldsymbol{e}_n)|}} (1 + o(1)) \text{ as } \beta \to \infty \tag{4.32}$$

As this formula involves g only through J, which is built using the gradient and Hessian of g, we can assume that g is a second order polynomial with respect to $\boldsymbol{x}$. Moreover, the hypothesis that $\boldsymbol{x}^* = \boldsymbol{e}_n$ leads to:

$$g(\boldsymbol{x}) = -\rho \boldsymbol{e}^t{}_n (\boldsymbol{x} - \boldsymbol{e}_n) - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{e}_n)^t \boldsymbol{M} (\boldsymbol{x} - \boldsymbol{e}_n)$$

where the symmetric matrix $\boldsymbol{M}$ writes:

$$\boldsymbol{M} = \left(\begin{array}{c|c} \boldsymbol{M}^* & 0 \\ \hline 0 & 0 \end{array}\right)$$

The main curvatures $(\kappa_i)_{i=1,\dots,n-1}$ of $\partial\mathcal{F}$ at $\boldsymbol{e}_n$ are exactly the eigenvalues of the Weingarten map $\mathcal{W}$ of g at $\boldsymbol{e}_n$ associated with eigenvectors that form an orthogonal basis of the tangent space of $\mathcal{F}$ at $\boldsymbol{e}_n$, see [Bre94, Definition 6]. The matrix $\boldsymbol{W}$ of $\mathcal{W}$ is given by:

$$\begin{aligned} \boldsymbol{W} &= -||\boldsymbol{\nabla} g(\boldsymbol{e}_n)||^{-1} \left(\boldsymbol{I}_n - \boldsymbol{\nabla} g(\boldsymbol{e}_n) \boldsymbol{\nabla}^t g(\boldsymbol{e}_n)\right) \boldsymbol{M} \\ &= -\rho^{-1} \left(\boldsymbol{I}_n - \rho^2 \boldsymbol{e}_n \boldsymbol{e}^t{}_n\right) \boldsymbol{M} \\ &= \left(\begin{array}{c|c} -\rho^{-1}\boldsymbol{M}^* & 0 \\ \hline 0 & 0 \end{array}\right) \end{aligned}$$

so the main curvatures $(\kappa_i)_{i=1,\dots,n-1}$ are the eigenvalues of $-\rho^{-1}\boldsymbol{M}^*$.

The gradient and Hessian of ℓ at $\boldsymbol{e}_n$ are:

$$\begin{aligned} \nabla \ell(\boldsymbol{e}_n) &= 2\beta^2 \frac{\phi'(\beta^2)}{\eta(\beta)} \boldsymbol{e}_n \\ \nabla^2 \ell(\boldsymbol{e}_n) &= \frac{2\beta^2}{\eta(\beta)} \left[2\beta^2 \phi'\left(\beta^2\right) \boldsymbol{e}_n \boldsymbol{e}^t{}_n + \phi'\left(\beta^2\right) \boldsymbol{I}_n\right] \end{aligned}$$

The gradient and the Hessian of g at $\boldsymbol{e}_n$ are:

$$\begin{aligned} \nabla g(\boldsymbol{e}_n) &= -\rho \boldsymbol{e}_n \\ \nabla^2 g(\boldsymbol{e}_n) &= -\boldsymbol{M} \end{aligned}$$

Taking into account that $\phi \leq 0$ and $\eta(\beta) \geq 0$ on $\mathcal{F}$, the matrix $\boldsymbol{H}(\boldsymbol{e}_n)$ reads:

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{e}_n) &= \frac{2\beta^2}{\eta(\beta)} \left[2\beta^2 \phi'\left(\beta^2\right) \boldsymbol{e}_n \boldsymbol{e}^t{}_n + \phi'\left(\beta^2\right) \boldsymbol{I}_n\right] - 2\beta^2 \frac{\phi'(\beta^2)}{\eta(\beta)} \rho^{-1} \boldsymbol{M} \\ &= \frac{2\beta^2}{\eta(\beta)} \left[2\beta^2 \phi'\left(\beta^2\right) \boldsymbol{e}_n \boldsymbol{e}^t{}_n + \phi'\left(\beta^2\right) \boldsymbol{I}_n - \phi'\left(\beta^2\right) \rho^{-1} \boldsymbol{M}\right] \end{aligned}$$

from which we get the expression of $J(\boldsymbol{e}_n)$:

$$J(\boldsymbol{e}_n) = \left(2\beta^2 \frac{\phi'(\beta^2)}{\eta(\beta)}\right)^2 \boldsymbol{e}^t{}_n \boldsymbol{C}(\boldsymbol{e}_n) \boldsymbol{e}_n = \left(2\beta^2 \frac{\phi'(\beta^2)}{\eta(\beta)}\right)^2 C_{nn}(\boldsymbol{e}_n)$$

where $C_{nn}(\boldsymbol{e}_n)$ is the cofactor of $H_{nn}(\boldsymbol{e}_n)$, i.e. the determinant of its $(n-1)\times(n-1)$ upper left block. We have:

$$\begin{aligned} C_{nn}(\boldsymbol{e}_n) &= \left(2\beta^2\frac{\phi'(\beta^2)}{\eta(\beta)}\right)^{n-1}\det\left(\boldsymbol{I}_{n-1}-\rho^{-1}\boldsymbol{M}\right) \\ &= \left(-2\beta^2\frac{\phi'(\beta^2)}{\eta(\beta)}\right)^{n-1}\prod_{i=1}^{n-1}(1-\kappa_i) \end{aligned}$$

so:

$$J(\boldsymbol{e}_n) = \left(-2\beta^2\frac{\phi'(\beta^2)}{\eta(\beta)}\right)^{n-1}\prod_{i=1}^{n+1}(1-\kappa_i) \tag{4.33}$$

Using (4.33) in (4.32), we get:

where $|\phi|$ has been substituted by $-\phi$ since $|\phi \leq 0$. Substituting ϕ and ϕ' by their expressions in terms of θ and θ' we get the expression (4.30), which is the Breitung approximation extended to a more general spherical case. □

In order to simplify the relation (4.30) we generalize Mill's ratio to the spherical cases we are interested in:

Proposition 4.25. *Let θ be the density generator of an n-dimensional spherical distribution satisfying the conditions of Proposition 4.24 and E be its one-dimensional marginal distribution function. Then we have the approximation:*

$$E(-\beta) = \left(-\frac{\pi\theta(\beta^2)}{\theta'(\beta^2)}\right)^{(n+1)/2}\frac{\theta(\beta^2)}{2\pi\beta}(1+o(1)) \quad \text{as } \beta\to\infty \tag{4.34}$$

Proof. When we write the relation (4.30) for a linear standard model G, we can evaluate exactly the integral defining $\mathbb{P}(\beta\mathcal{F})$ to get $\mathbb{P}(\beta\mathcal{F}) = E(-\beta)$. As all the curvatures are equal to zero, we get $E(-\beta) = \left(-\frac{\pi\theta(\beta^2)}{\theta'(\beta^2)}\right)^{(n+1)/2}\frac{\theta(\beta^2)}{2\pi\beta}(1+o(1))$ as $\beta\to\infty$. □

Using Propositions 4.24 and 4.25, we get the final form of Breitung's approximation extended to a more general spherical case:

Theorem 4.26. *Let G be a standard model and $\boldsymbol{U}$ a random vector with a spherical distribution of density generator θ such that the conditions of Theorem 4.23 are fulfilled. Then, the asymptotic relation (4.30) reads:*

$$\mathbb{P}(\beta\mathcal{F}) = E(-\beta)\left(\prod_{i=1}^{n-1}\frac{1}{\sqrt{1-\kappa_i}}\right)(1+o(1)) \quad \text{as } \beta\to\infty \tag{4.35}$$

where $\boldsymbol{x}^* = \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{F}}||\boldsymbol{x}||$, $(\kappa_i)_{i=1,\ldots,n-1}$ are the main curvatures of $\partial\mathcal{F}$ at $\boldsymbol{x}^*$ and θ' is the first derivative of θ.

Remark 4.27. In practice, if one knows that the standard model G is in the parametric family $\mathcal{G}_\beta$, then the value of β as well as the curvatures $(\kappa_i)_{i=1,\dots,n-1}$ can be recovered directly from the design point $\boldsymbol{u}^*$ and the curvatures associated with G:

$$\begin{aligned}\beta =& \|\boldsymbol{u}\|* \\ \kappa_i =& -\beta\kappa_i^G \quad \forall i \in \{1,\dots,n-1\}\end{aligned}$$

where $\left(\kappa_i^G\right)_{i=1,\dots,n-1}$ are the main curvatures of $\partial\mathcal{D}$ at $\boldsymbol{u}^*$, $\partial\mathcal{D}$ and $\partial\mathcal{F}$ having opposite orientation convention.

Remark 4.28. The monotonicity condition and the separability condition (4.29) made on ϕ are not really mandatory. They could be relaxed, for instance by imposing them only asymptotically. It would gives the result (4.35) for a broader class of spherical distributions.

4.5 Conclusion

In this chapter, we have given an introduction to elliptical distributions and copulas, and we proposed a generalization of the Nataf transformation to any distribution with elliptical copula. In order to make an effective use of this generalized transformation, we derived the associated FORM and SORM/Breitung approximation of the probability of failure. These approximations appear to be very natural extensions of the normal copula case associated with the usual Nataf transformation.

Chapter 5

Do Rosenblatt and Nataf isoprobabilistic transformations really differ?

In this chapter, we explore the relationship between two isoprobabilistic transformations widely used in the community of reliability analysts, namely the generalized Nataf transformation and the Rosenblatt transformation.

The main results of this chapter are the demonstration that the Rosenblatt transformation using the canonical order of conditioning is identical to the generalized Nataf transformation in the normal copula case, which is the most usual case in reliability analysis since it corresponds to the classical Nataf transformation. Then, we show that it is not possible to extend the Rosenblatt transformation to distributions with general elliptical copula the way the Nataf transformation has been generalized. Finally, we explore the effect of the conditioning order of the Rosenblatt transformation on the usual reliability indicators obtained from a FORM or SORM method. We show that in the normal copula case, all these reliability indicators, excepted the importance factors, are unchanged whatever the conditioning order one choose.

These results are illustrated with two numerical applications that illustrate the previous results.

This work has been published in [LD09a].

5.1 Introduction

We presented in the previous chapter the FORM and SORM approximations to compute probabilities of failure as an alternative to the simulation methods. These methods are based on an isoprobabilistic transformation that maps the physical space into a new space called the **standard space**. To this end, two isoprobabilistic transformations are presented in the literature: the generalized Nataf transformation that has been presented in details in the previous chapter, and the Rosenblatt transformation [Ros52].

The main objective of this chapter is to compare the generalized Nataf transformation with the Rosenblatt one and to prove that they are identical in the normal copula case, which is the most common case in actual reliability studies as it corresponds to the use of

the classical Nataf transformation.

We also study the possibility to modify the Rosenblatt transformation in a similar way the Nataf transformation has been extended to lead to a non-normal standard space.

The second objective of this chapter is to study the impact of the conditioning order of the Rosenblatt transformation on the usual reliability indicators obtained after an analytical FORM / SORM method, with a focus on the normal copula case.

We denote by $C^N_{\boldsymbol{R}}$ a normal copula whose correlation matrix is $\boldsymbol{R}$. We suppose that $\boldsymbol{R}$ is a symmetric positive definite matrix.

We denote by $\mathbb{M}_{n,n}(\mathbb{R})$ the algebra of real square matrices of dimension n, by $\mathcal{O}_n(\mathbb{R})$ the multiplicative sub-group of orthogonal matrices and by $\mathcal{GL}_n(\mathbb{R})$ the multiplicative sub-group of invertible matrices.

If $\boldsymbol{R} = (r_{ij})_{1\leq i,j\leq n} \in \mathbb{M}_{n,n}(\mathbb{R})$, then $\boldsymbol{R}_k$ is its k-leading sub-block:

$$\boldsymbol{R}_k = (r_{ij})_{1\leq i,j\leq k} \tag{5.1}$$

and $\boldsymbol{R}^k$ is the $(k+1)$-th partial column vector:

$$\boldsymbol{R}^k = (r_{1,k+1}, \ldots, r_{k,k+1})^t \tag{5.2}$$

We call **standard space** the image space of an isoprobabilistic transformation.

5.2 The generalized Nataf and Rosenblatt transformations

We presented the generalized Nataf transformation in the previous chapter, with a decomposition in three steps $T = T_2 \circ T_1''^{gen} \circ T_1'$. Here, we contract the action of $T_1''^{gen}$ and T_1' into a unique transformation $T_1^{gen} = T_1''^{gen} \circ T_1'$ to get the representation that is most adapted to a comparison with the Rosenblatt transformation. We also add an explicit mention to the Nataf name in order to avoid confusion with the Rosenblatt transformation: $T^{gen}_{Nataf} = T_2 \circ T_1^{gen}$.

The interest of this transformation for uncertainty quantification purpose is that it maps random vectors with elliptical copulas into random vectors with spherical distribution, which is mandatory to use approximation methods such as the FORM or SORM approximations. More precisely, if the random vector $\boldsymbol{X}$ has a joint distribution function with marginal distribution functions $F_1, \ldots, F_n$ and copula $C_{\psi,\boldsymbol{R}}$, then the random vector $\boldsymbol{U} = T(\boldsymbol{X})$ is distributed according to the standard spherical representative distribution $E_{\boldsymbol{0},\boldsymbol{1},\boldsymbol{I}_n,\psi}$ of $C_{\psi,\boldsymbol{R}}$.

Another widely used isoprobabilistic transformation is the Rosenblatt transformation [Ros52], defined as follows:

Definition 5.1. Let F be a continuous n-dimensional distribution function. The **Rosenblatt transformation** $T^R : \mathbb{R}^n \to \mathbb{R}^n$ associated with F is defined by:

$$T^R = T_2^R \circ T_1^R \tag{5.3}$$

where the transformations T_1^R and T_2^R are given by:

$$T_1^R : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \\ \boldsymbol{x} & \mapsto & \boldsymbol{w} = \begin{pmatrix} F_1(x_1) \\ \vdots \\ F_{k|1,\dots,k-1}(x_k|x_1,\dots,x_{k-1}) \\ \vdots \\ F_{n|1,\dots,n-1}(x_n|x_1,\dots,x_{n-1}) \end{pmatrix} \end{array} \right. \tag{5.4}$$

$$T_2^R : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \\ \boldsymbol{w} & \mapsto & \boldsymbol{u} = \begin{pmatrix} \Phi^{-1}(w_1) \\ \vdots \\ \Phi^{-1}(w_n) \end{pmatrix} \end{array} \right.$$

where $F_{k|1,\dots,k-1}$ is the conditional distribution function defined in (1.1).

The interest of this transformation is in its action on a random vector $\boldsymbol{X}$ with joint distribution F, as given in the following theorem (see [DM05, Chapter 7.2] or [Ros52]):

Theorem 5.2. *Let $\boldsymbol{X}$ be a n-dimensional random vector with continuous joint distribution function F and T^R the associated Rosenblatt transformation. Then $\boldsymbol{U} = T^R\boldsymbol{X}$ is distributed according to the standard n-dimensional normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$.*

Proof. Let $\boldsymbol{Z}$ be the random vector defined by $\boldsymbol{Z} = T_1^R(\boldsymbol{X})$. By construction, $\boldsymbol{Z}$ takes its values in $[0,1]^n$. To compute the distribution of $\boldsymbol{Z}$, let g be a real valued bounded continuous function defined on $[0,1]^n$. We have:

$$\begin{aligned} \mathbb{E}\left[g(\boldsymbol{Z})\right] =& \mathbb{E}\left[g(T_1^R(\boldsymbol{X}))\right] \\ =& \int_{\mathbb{R}^n} g(T_1^R(\boldsymbol{x}))p_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ =& \int_{\Omega} g(T_1^R(\boldsymbol{x}))p_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \end{aligned}$$

where $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n \,\|\, p_{\boldsymbol{X}}(\boldsymbol{x}) > 0\}$. The change of variable $\boldsymbol{z} = T_1^R(\boldsymbol{x})$ is a diffeomorphism between Ω and $(0,1)^n$, and we have $\mathrm{d}\boldsymbol{z} = |\det(\boldsymbol{J}(\boldsymbol{x}))|\,\mathrm{d}\boldsymbol{x}$, where $\boldsymbol{J}$ is the Jacobian matrix of T_1^R defined by:

$$\begin{aligned} J_{i,j} =& \frac{\partial (T_1^R)_i(\boldsymbol{x})}{\partial x_j} \\ =& \frac{\partial F_{i|1,\dots,i-1}(x_i|x_1,\dots,x_{i-1})}{\partial x_j} \end{aligned}$$

We note that for $j > i$, $J_{i,j} = 0$ as $F_{i|1,\dots,i-1}$ does not depend on x_j. The matrix $\boldsymbol{J}$ is then lower triangular, and its determinant is the product of its diagonal elements:

$$\det(\boldsymbol{J}(\boldsymbol{x})) = \prod_{k=1}^{n} \frac{\partial F_{k|1,\dots,k-1}(x_k|x_1,\dots,x_{k-1})}{\partial x_k} = \prod_{k=1}^{n} p_{k|1,\dots,k-1}(x_k|x_1,\dots,x_{k-1}) = p_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$$

so $\mathrm{d}\boldsymbol{z} = p_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ and we get:

$$\begin{aligned}\mathbb{E}\left[g(\boldsymbol{Z})\right] &= \int_{[0,1]^n} g(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} \\ &= \int_{\mathbb{R}^n} g(\boldsymbol{z})\mathbb{1}_{[0,1]^n}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}\end{aligned}$$

which proves that $\boldsymbol{Z}$ is uniformly distributed over $[0,1]^n$, i.e it has independent components which are all uniformly distributed over $[0,1]$. By Proposition 1.18, the transformation T_2^R maps $\boldsymbol{Z}$ into a random vector $\boldsymbol{Y} = T_2^R(\boldsymbol{Z})$ with the same copula, so $\boldsymbol{Y}$ has independent components. Its marginal distribution functions are all equal to Φ thus $\boldsymbol{Y}$ has a standard normal distribution. □

In order to ease the further comparison between the generalized Nataf transformation and the Rosenblatt one, it is useful to rewrite the Rosenblatt transformation as follows:

Proposition 5.3. *Let F be a n-dimensional continuous distribution function with univariate marginal distribution functions $F_1, \ldots, F_n$ and copula C. The* **new formulation of the Rosenblatt transformation** *T^{NR} associated with F is defined by:*

$$T^{NR} = T^R \circ T_0 \tag{5.5}$$

where T_0 is given by :

$$T_0 : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \\ \boldsymbol{x} & \mapsto & \boldsymbol{w} = \begin{pmatrix} \Phi^{-1} \circ F_1(x_1) \\ \vdots \\ \Phi^{-1} \circ F_n(x_n) \end{pmatrix} \end{array} \right. \tag{5.6}$$

where T^R the Rosenblatt transformation of Definition 5.1 with the distribution function of $T_0(\boldsymbol{X})$, i.e a distribution with standard normal marginal distributions and copula C.

Proof. Let us note that $\boldsymbol{u}^{NR} = T^{NR}(\boldsymbol{x}) = T_2^R \circ T_1^R \circ T_0(\boldsymbol{x})$.

If $\boldsymbol{w} = T_0(\boldsymbol{x})$, then, thanks to (5.4), the k^{th} component of $\boldsymbol{u}^{NR}$ writes:

$$u_k^{NR} = \Phi^{-1} \circ G_{k|1,\ldots,k-1}(w_k|w_1, \ldots, w_{k-1}) \tag{5.7}$$

where $G_{k|1,\ldots,k-1}(w_k|w_1, \ldots, w_{k-1})$ is the conditional distribution function of a random vector $\boldsymbol{W}$ distributed as $T_0(\boldsymbol{X})$, where $\boldsymbol{X}$ is a random vector with marginal distribution functions $F_1, \ldots, F_n$ and copula C. We note by $C^{\boldsymbol{W}}$ the copula of $\boldsymbol{W}$. Thanks to Proposition 1.13, the distribution function $G_{k|1,\ldots,k-1}(w_k|w_1, \ldots, w_{k-1})$ writes:

$$G_{k|1,\ldots,k-1}(w_k|w_1, \ldots, w_{k-1}) = C^{\boldsymbol{W}}_{k|1,\ldots,k-1}\left(G_k(w_k)|G_1(w_1), \ldots, G_{k-1}(w_{k-1})\right) \tag{5.8}$$

From Proposition 1.18, it follows that $\boldsymbol{X}$ and $\boldsymbol{W}$ have the same copula C. Furthermore, by construction of $\boldsymbol{W}$, we have $G_k = \Phi$ and $\Phi(w_k) = F_k(x_k)$. Then, relation (5.8) rewrites:

$$G_{k|1,\ldots,k-1}(w_k|w_1, \ldots, w_{k-1}) = C_{k|1,\ldots,k-1}\left(F_k(x_k)|F_1(x_1), \ldots, F_{k-1}(x_{k-1})\right) \tag{5.9}$$

which finally leads to the relation:

$$G_{k|1,\ldots,k-1}(w_k|w_1, \ldots, w_{k-1}) = F_{k|1,\ldots,k-1}(x_k|x_1, \ldots, x_{k-1}) \tag{5.10}$$

and then to:

$$T_k^{NR}(\boldsymbol{x}) = \Phi^{-1} \circ F_{k|1,\ldots,k-1}(x_k|x_1, \ldots, x_{k-1}) = T_k^R(\boldsymbol{x}) \tag{5.11}$$

where T^R is the classical Rosenblatt transformation associated to F. □

5.3 Do generalized Nataf and Rosenblatt transformations really differ?

In this section, we first consider the case where the copula of $\boldsymbol{X}$ is normal, which is the most usual case in reliability analysis since it corresponds to the case where the classical Nataf transformation applies.
Then, we make the comparison in all the other cases: non-normal elliptical copulas and non-elliptical copulas.

5.3.1 The normal copula case

The new formulation (5.5) of the Rosenblatt transformation makes it easier to show that when $\boldsymbol{X}$ has a normal copula, both transformations are identical:

Proposition 5.4. *Let $\boldsymbol{X}$ in $\mathbb{R}^n$ be a continuous random vector defined by its univariate marginal distribution functions F_i and its copula $C_{\boldsymbol{R}}^{\mathcal{N}}$ supposed to be normal, with non-singular correlation matrix $\boldsymbol{R}$. Then, the Rosenblatt transformation and the generalized Nataf one are identical:*

$$T^R(\boldsymbol{X}) = T^{GN}(\boldsymbol{X}) \tag{5.12}$$

We recall without demonstration the conditional expectation and covariance matrix of a normal random vector, the action of an affine transformation on a normal random vector and an elementary result on orthogonal matrices that will be used in the demonstration of Proposition 5.4.

Proposition 5.5. *Let $\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2)$ be a $n_1 + n_2$-dimensional normal random vector such that $\mathbf{Cov}\left[\boldsymbol{U_1}, \boldsymbol{U_1}\right]$ is nonsingular. Then the conditional distribution of $\boldsymbol{U}_2$ given $\boldsymbol{U}_1$ is the normal distribution with mean vector and covariance matrix defined by:*

$$\begin{cases} \mathbb{E}\left[\boldsymbol{U}_2 \,|\, \boldsymbol{U}_1\right] = \mathbb{E}\left[\boldsymbol{U}_2\right] + \mathbf{Cov}\left[\boldsymbol{U_2}, \boldsymbol{U_1}\right]\left[\mathbf{Cov}\left[\boldsymbol{U_1}, \boldsymbol{U_1}\right]\right]^{-1}\left(\boldsymbol{U}_1 - \mathbb{E}\left[\boldsymbol{U}_1\right]\right) \\ \mathbf{Cov}\left[\boldsymbol{U_2} \,|\, \boldsymbol{U_1}\right] = \mathbf{Cov}\left[\boldsymbol{U_2}, \boldsymbol{U_2}\right] - \mathbf{Cov}\left[\boldsymbol{U_2}, \boldsymbol{U_1}\right]\left[\mathbf{Cov}\left[\boldsymbol{U_1}, \boldsymbol{U_1}\right]\right]^{-1}\mathbf{Cov}\left[\boldsymbol{U_1}, \boldsymbol{U_2}\right] \end{cases} \tag{5.13}$$

Proposition 5.6. *Let $\boldsymbol{U}$ in $\mathbb{R}^n$ be a normal vector, with mean vector is $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{A}$ a deterministic matrix in $\mathbb{M}_{n,p}(\mathbb{R})$ and $\boldsymbol{b}$ in $\mathbb{R}^p$ a deterministic vector. Then $\boldsymbol{V} = \boldsymbol{A}\,\boldsymbol{X} + \boldsymbol{b}$ is a normal vector which mean vector and covariance matrix are defined by:*

$$\begin{cases} \mathbb{E}\left[\boldsymbol{V}\right] &= \boldsymbol{A}\,\boldsymbol{\mu} + \boldsymbol{b} \\ \mathbf{Cov}\left[\boldsymbol{V}\right] &= \boldsymbol{A}\,\boldsymbol{\Sigma}\,\boldsymbol{A}^t \end{cases} \tag{5.14}$$

Proposition 5.7. *Let $\mathcal{T}^+(\mathbb{R})$ be the set of lower triangular matrix of $\mathbb{M}_{n,n}(\mathbb{R})$ with positive diagonal elements. Then $\mathcal{T}^+(\mathbb{R})$ is a multiplicative subgroup of $\mathcal{GL}_n(\mathbb{R})$.*
Furthermore, $\mathcal{T}^+(\mathbb{R}) \cap \mathcal{O}_n(\mathbb{R}) = \{\boldsymbol{I}_n\}$.

Proof. We can now start to demonstrate Proposition 5.4, using the new formulation of the Rosenblatt transformation of Proposition 5.3, whose different steps are the following ones:

$$T^{NR} : \boldsymbol{X} \overset{T_0}{\longmapsto} \boldsymbol{W} \overset{T_1^R}{\longmapsto} \boldsymbol{Y} \overset{T_2^R}{\longmapsto} \boldsymbol{U} \tag{5.15}$$

Let us note $\boldsymbol{S}_{k-1} = (W_1, \ldots, W_{k-1})^t$ and V_k a random variable distributed as W_k given $\boldsymbol{S}_{k-1}$.

As $\boldsymbol{X}$ has a normal copula, $\boldsymbol{W}$ is a n-dimensional normal vector whose univariate marginal distributions are standard normal and whose correlation matrix is $\boldsymbol{R}$.

Proposition 5.5 gives that for all k, V_k follows a univariate normal distribution and relation (5.13) leads to:

$$\begin{array}{rcl} \mathbb{E}\left[V_k\right] & = & \mathbb{E}\left[W_k\right] + \mathbf{Cov}\left[\boldsymbol{W_k}, \boldsymbol{S_{k-1}}\right]\left[\mathbf{Cov}\left[\boldsymbol{S_{k-1}}, \boldsymbol{S_{k-1}}\right]\right]^{-1}\left(\boldsymbol{S}_{k-1} - \mathbb{E}\left[\boldsymbol{S}_{k-1}\right]\right) \\ & = & \mathbf{Cov}\left[\boldsymbol{W_k}, \boldsymbol{S_{k-1}}\right]\left[\mathbf{Cov}\left[\boldsymbol{S_{k-1}}, \boldsymbol{S_{k-1}}\right]\right]^{-1}\boldsymbol{S}_{k-1} \end{array} \tag{5.16}$$

the matrix $\mathbf{Cov}\left[\boldsymbol{S_{k-1}}, \boldsymbol{S_{k-1}}\right]$ being nonsingular as it is the $(k-1)\times(k-1)$ upper left square block of the nonsingular matrix $\boldsymbol{R}$.

We have $\mathbf{Cov}\left[\boldsymbol{W}\right] = \mathbf{Cor}\left[\boldsymbol{W}\right] = \boldsymbol{R}$, and given the notations (5.1) and (5.2), we have:

$$\mathbb{E}\left[V_k\right] = \left(\boldsymbol{R}^{k-1}\right)^t \left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{S}_{k-1} \tag{5.17}$$

Furthermore, relation (5.13) also leads to:

$$\begin{array}{rcl} \mathbf{Var}\left[V_k\right] & = & \mathbf{Var}\left[W_k\right] - \mathbf{Cov}\left[\boldsymbol{W_k}, \boldsymbol{S_{k-1}}\right]\left[\mathbf{Cov}\left[\boldsymbol{S_{k-1}}, \boldsymbol{S_{k-1}}\right]\right]^{-1}\mathbf{Cov}\left[\boldsymbol{S_{k-1}}, \boldsymbol{W_k}\right] \\ & = & 1 - \left(\boldsymbol{R}^{k-1}\right)^t \left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{R}^{k-1} \end{array} \tag{5.18}$$

Given relations (5.17) and (5.18), the k^{th} component of $\boldsymbol{Y}$ is defined by:

$$Y_k = F^{\boldsymbol{W}}_{k|1,\ldots,k-1}(W_k|W_1,\ldots,W_{k-1}) = \Phi\left(\frac{W_k - \left(\boldsymbol{R}^{k-1}\right)^t\left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{S}_{k-1}}{\sqrt{1 - \left(\boldsymbol{R}^{k-1}\right)^t\left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{R}^{k-1}}}\right) \tag{5.19}$$

Finally, we obtain:

$$U_k = \Phi^{-1}(Y_k) = \frac{W_k - \left(\boldsymbol{R}^{k-1}\right)^t\left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{S}_{k-1}}{\sqrt{1 - \left(\boldsymbol{R}^{k-1}\right)^t\left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{R}^{k-1}}} = \boldsymbol{A}_k\boldsymbol{W} \tag{5.20}$$

where for all $k \in [1,n]$, $\boldsymbol{A}_k = (a_{k,1},\ldots,a_{k,k},0,\ldots,0) \in \mathcal{M}_{1n}(\mathbb{R})$ with:

$$\left\{\begin{array}{rcl} a_{k,k} & = & \left[\sqrt{1 - \left(\boldsymbol{R}^{k-1}\right)^t\left[\boldsymbol{R}_{k-1}\right]^{-1}\boldsymbol{R}^{k-1}}\right]^{-1} \\ a_{k,j} & = & -a_{k,k}\displaystyle\sum_{i=1}^{k-1} r_{1i}r_{ji} \quad \text{for } \forall j \in [1,k-1] \end{array}\right. \tag{5.21}$$

As $\boldsymbol{A}_k$ is a row matrix, U_k only depends on $\boldsymbol{S_k}$. Let $\tilde{\boldsymbol{\Gamma}}$ be the lower triangular matrix whose kth row is $\boldsymbol{A}_k$. Then relation (5.20) implies that:

$$\boldsymbol{U} = \tilde{\boldsymbol{\Gamma}}\,\boldsymbol{W} \tag{5.22}$$

which is very close to relation (3.3). It remains to show that $\tilde{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}$.

Proposition 5.6 implies that $\mathbf{Cov}\left[\boldsymbol{U}\right] = \tilde{\boldsymbol{\Gamma}}\,\boldsymbol{R}\,\tilde{\boldsymbol{\Gamma}}^t$ and $\mathbf{Cov}\left[\boldsymbol{U}\right] = \boldsymbol{I}_n$ by construction of $\boldsymbol{U}$. If $\boldsymbol{L}$ is the Cholesky factor of $\boldsymbol{R}$, then $\boldsymbol{R} = \boldsymbol{L}\,\boldsymbol{L}^t$, and $(\tilde{\boldsymbol{\Gamma}}\,\boldsymbol{L})(\tilde{\boldsymbol{\Gamma}}\,\boldsymbol{L})^t = \boldsymbol{I}_n$, which leads to $\tilde{\boldsymbol{\Gamma}}\,\boldsymbol{L} \in \mathcal{O}_n(\mathbb{R})$.
Furthermore, by construction, $\tilde{\boldsymbol{\Gamma}} \in \mathcal{T}^+(\mathbb{R})$. As $\boldsymbol{L} \in \mathcal{T}^+(\mathbb{R})$, Proposition 5.7 implies that $\tilde{\boldsymbol{\Gamma}}\,\boldsymbol{L} \in \mathcal{T}^+(\mathbb{R})$ and $\tilde{\boldsymbol{\Gamma}}\,\boldsymbol{L} = \boldsymbol{I}_n$, which rewrites $\tilde{\boldsymbol{\Gamma}} = \boldsymbol{L}^{-1} = \boldsymbol{\Gamma}$.

In conclusion, we showed that in the case where $\boldsymbol{X}$ has a normal copula, we have the relation $T_2^R \circ T_1^R \circ T_0(\boldsymbol{X}) = T_2^N \circ T_1^N(\boldsymbol{X})$ which leads to :

$$T^R(\boldsymbol{X}) = T^N(\boldsymbol{X}) \text{ a.s} \tag{5.23}$$

Thus, the equivalence of the Rosenblatt transformation and the generalized Nataf transformation in the normal copula case is shown. □

5.3.2 The other cases

In the case where the copula of $\boldsymbol{X}$ is elliptical but non-normal, both isoprobabilistic transformations differ as their associated standard spaces are different. As a matter of fact, the standard spaces of the generalized Nataf is associated with the standard spherical representative of the elliptical family that defines the elliptical copula, whereas the standard space of the Rosenblatt transformation is associated to the normal distribution.

At this step, it is interesting to check whether it is possible to modify the Rosenblatt transformation in order to make its standard space be the same as the one associated with the generalized Nataf transformation.

In the previous chapter, we have recalled that the essential characteristic of the standard space is the spherical symmetry of its associated distribution, which gives a sense to the FORM and SORM approximations of the event probability.
Let us note that by construction, because of the conditioning step T_1^R, the Rosenblatt transformation leads to a final vector $\boldsymbol{U}$ with an independent copula.

Proposition 5.8. *The only spherical distributions with independent components are the normal distributions with zero mean and scalar covariance matrix proportional to the identity.*

See [AL82] for a demonstration.

Thus, the only way to map a random vector with independent copula into a random vector following a spherical distribution, is to map it into a normal vector such as described in this proposition: thus, the standard space of the Rosenblatt transformation is necessarily the normal one.
Therefore, the standard space of the Rosenblatt transformation and the standard space of the generalized Nataf transformation only coincide in the normal copula case.

Finally, for all the other cases where the copula of $\boldsymbol{X}$ is not elliptical, the generalized Nataf transformation is not defined and the comparison with the Rosenblatt transformation not possible.

5.4 Impact of the conditioning order in the Rosenblatt transformation in the normal copula case

In the literature [DM05], the presentation of the Rosenblatt transformation is given with the warning that the conditioning order in step T_1^R has an impact on the results obtained from a FORM/SORM method. This warning was already present in the seminal paper [Ros52] of Rosenblatt.
Let us call **canonical order** the order presented in the relation (5.4).

In that section, we study the impact of a change in the conditioning order of the Rosenblatt transformation on the quantities evaluated in the context of the use of the FORM or SORM methods : the design point, which is used through its norm (reliability index) and its components for the computation of the importance factors, and the curvatures of the limit state surface at the design point in the standard space, where the limit state surface is the boundary of the subspace of parameters verifying the event (for SORM approximation).

In the case where the copula of $\boldsymbol{X}$ is not normal, it has already been shown that such a change has an impact on all these elements : see the example quoted by [Dol83] and discussed by several authors, for example [DM05] and [Lem05].

However, this is not always the case. We will study in more detail the most frequent situation where the copula of $\boldsymbol{X}$ is normal since, as mentioned previously, it is the copula induced by the traditional use of the classical Nataf transformation.

Let us suppose now that we change the order of conditioning. It is equivalent to consider the introduction of a new step in the Rosenblatt transformation between the steps T_0 and T_1^R of relation (5.15) in order to make a permutation $\boldsymbol{P} \in \mathfrak{S}_n(\mathbb{R})$ of the components of $\boldsymbol{W}$ to get $\boldsymbol{W}_2$. The Rosenblatt transformations using the canonical order or an arbitrary order are summarized graphically in Figure 5.1.

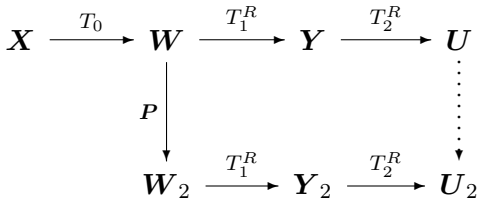


Figure 5.1: Rosenblatt transformations when the conditioning of the components W_k follow the canonical order or an arbitrary order.

We have the following result:

Proposition 5.9. *In the normal copula case, changing the order of the conditioning in the Rosenblatt transformation consists in making an orthogonal transformation in the standard space of the Rosenblatt transformation.*
More precisely, if we note $\boldsymbol{P} \in \mathfrak{S}_n(\mathbb{R})$ the permutation matrix associated to the arbitrary order, T_2^R the Rosenblatt transformation associated to this ordering, $\boldsymbol{U}_2 = T_2^R(\boldsymbol{X})$ and

$\boldsymbol{U} = T^R(\boldsymbol{X})$, then we have :

$$\exists \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R}) \,/\, \boldsymbol{U}_2 = \boldsymbol{Q}\boldsymbol{U} \tag{5.24}$$

where $\boldsymbol{Q}$ and $\boldsymbol{P}$ are in the same connected component of $\mathcal{O}_n(\mathbb{R})$, it means $\det(\boldsymbol{P}) = \det(\boldsymbol{Q}) = \pm 1$.

According to the notations of Figure 5.1, if $\boldsymbol{R}$ is the correlation matrix of the normal copula of $\boldsymbol{X}$, $\boldsymbol{R}_2$ the one of $\boldsymbol{W}_2$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_2$ the inverse of their respective Cholesky factors, then the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ are linked by:

$$\boldsymbol{Q} = \boldsymbol{\Gamma}_2 \, \boldsymbol{P} \, \boldsymbol{\Gamma}^{-1} \tag{5.25}$$

and in general, neither $\boldsymbol{Q} = \boldsymbol{P}$ nor $\boldsymbol{Q} \in \mathfrak{S}_n(\mathbb{R})$.

The following result will help for the demonstration of Proposition 5.9:

Proposition 5.10. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two deterministic matrices in $\mathbb{M}_{n,n}(\mathbb{R})$, with $\boldsymbol{B}$ invertible. Then we have:*

$$\boldsymbol{A}\,\boldsymbol{A}^t = \boldsymbol{B}\,\boldsymbol{B}^t \Longrightarrow \boldsymbol{B}^{-1}\,\boldsymbol{A} \in \mathcal{O}_n(\mathbb{R}) \tag{5.26}$$

which means that $\exists \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})$ such that $\boldsymbol{A} = \boldsymbol{B}\,\boldsymbol{Q}$.

Proof. As a matter of fact, we have the following implications:

$$(\boldsymbol{B}^{-1}\,\boldsymbol{A})(\boldsymbol{B}^{-1}\,\boldsymbol{A})^t = \boldsymbol{B}^{-1}\,\boldsymbol{A}\,\boldsymbol{A}^t\,\boldsymbol{B}^{-t} = \boldsymbol{B}^{-1}\,\boldsymbol{B}\,\boldsymbol{B}^t\,\boldsymbol{B}^{-t} = \boldsymbol{I}_n \tag{5.27}$$

which leads to the result of Proposition 5.10. □

Proof. We are now ready to prove Proposition 5.9. As $\boldsymbol{W}_2 = \boldsymbol{P}\,\boldsymbol{W}$, $\boldsymbol{W}_2$ is a normal vector which correlation matrix verifies $\boldsymbol{R}_2 = \boldsymbol{P}\,\boldsymbol{R}\,\boldsymbol{P}^t$ and whose Cholesky factor is $\boldsymbol{L}_2 = \boldsymbol{\Gamma}_2^{-1}$. Therefore, $\boldsymbol{R}_2 = \boldsymbol{L}_2\,\boldsymbol{L}_2{}^t = (\boldsymbol{P}\,\boldsymbol{L})(\boldsymbol{P}\,\boldsymbol{L})^t$. Proposition 5.10 leads to:

$$\exists \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R}) \text{ such that } \boldsymbol{P}\,\boldsymbol{L} = \boldsymbol{L}_2\,\boldsymbol{Q} \tag{5.28}$$

By multiplying the relation (5.28) on the left by $\boldsymbol{\Gamma}_2$ and on the right by $\boldsymbol{\Gamma}$, it rewrites:

$$\boldsymbol{\Gamma}_2\,\boldsymbol{P} = \boldsymbol{Q}\,\boldsymbol{\Gamma} \tag{5.29}$$

which leads to the relation between $\boldsymbol{P}$ and $\boldsymbol{Q}$ given in relation (5.25).

We showed that in the normal copula case, the mapping from $\boldsymbol{W}_2$ into $\boldsymbol{U}_2$ is linear such as: $\boldsymbol{U}_2 = \boldsymbol{\Gamma}_2\,\boldsymbol{W}_2$. Finally, we obtain:

$$\boldsymbol{U}_2 = \boldsymbol{\Gamma}_2\,\boldsymbol{P}\,\boldsymbol{W} \tag{5.30}$$

Relations (5.29) and (5.30) finally imply that:

$$\boldsymbol{U}_2 = \boldsymbol{Q}\,\boldsymbol{\Gamma}\,\boldsymbol{W} = \boldsymbol{Q}\,\boldsymbol{U} \tag{5.31}$$

as required.

Given that $\det(\boldsymbol{\Gamma}) > 0$ and $\det(\boldsymbol{\Gamma}_2) > 0$, relation (5.25) implies that $\det(\boldsymbol{Q})$ and $\det(\boldsymbol{P})$ have the same sign, which means that they belong to the same connected component of $\mathcal{O}_n(\mathbb{R})$. □

In conclusion, if the random vector $\boldsymbol{X}$ has a normal copula, the effect of changing the order of conditioning in the Rosenblatt transformation with respect to the canonical order is to apply a further orthogonal transformation after applying the Rosenblatt transformation associated to the canonical ordering. It changes the location of the design point, i.e its components, but neither its norm nor the curvatures of the limit state surface at the design point.
Thus, in the context of the FORM or SORM method, the following quantities do not depend on the conditioning order of the Rosenblatt transformation :

- The Hasofer reliability index [HL74], which is the norm of the design point,
- The FORM approximation of the event probability which relies only on the Hasofer reliability index,
- The several SORM approximations of the event probability which rely on both the Hasofer reliability index and the curvatures of the limit state function at the design point.

However, the importance factors change in a way which is not in general a permutation of the values obtained using the canonical order: relation (5.25) implies that in general, $\boldsymbol{Q} \neq \boldsymbol{P}$.

To be more precise, we have the following result:

Proposition 5.11.
The random vector $\boldsymbol{X}$ has an independent copula if and only if for all permutation matrix $\boldsymbol{P}$, $\boldsymbol{Q} = \boldsymbol{P}$.

Proof. The first implication is obvious: if $\boldsymbol{X}$ has an independent copula, the correlation matrix $\boldsymbol{R}$ is equal to the identity matrix $\boldsymbol{I}_n$, which implies that $\boldsymbol{R}_2 = \boldsymbol{I}_n$, $\boldsymbol{\Gamma} = \boldsymbol{I}_n$, $\boldsymbol{\Gamma}_2 = \boldsymbol{I}_n$ and finally $\boldsymbol{Q} = \boldsymbol{P}$.
The second implication derives from the following computation. By definition of $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\Gamma}$, we have:

$$\boldsymbol{Q} = \boldsymbol{P} \Longrightarrow \boldsymbol{L}_2 = \boldsymbol{P}\,\boldsymbol{L}\,\boldsymbol{P}^t \tag{5.32}$$

which implies the following relation on the coefficients of $\boldsymbol{L}_2 = (\ell^2_{i,j})_{1\leq i,j\leq n}$ and $\boldsymbol{L} = (\ell_{i,j})_{1\leq i,j\leq n}$:

$$\ell^2_{i,j} = \ell_{\sigma(i),\sigma(j)} \tag{5.33}$$

where σ is the permutation associated to $\boldsymbol{P}$.
Thus, given that $\boldsymbol{L}$ and $\boldsymbol{L}_2$ are lower triangular matrices, if the relation (5.33) must hold for all the permutations σ, it must hold in particular for any transposition τ_{ij} that exchanges i and j, thus if $i < j$, $\ell^2_{i,j} = 0$ by construction, thus $\ell_{\sigma(i),\sigma(j)} = \ell_{ji} = 0$: $\boldsymbol{L}$ is a diagonal matrix and consequently, $\boldsymbol{R} = \boldsymbol{I}_n$, which is equivalent to the independence of the components of $\boldsymbol{X}$ in the normal copula case. □

In conclusion, a permutation with respect to the canonical order on the components of $\boldsymbol{X}$ always corresponds to the same permutation with respect to the canonical order of the components of the standard space random vector only in the independent case. Otherwise, the choice of the conditioning order does not translate into a simple permutation of the values of the design point coordinates.

The lack of invariance of the importance factors with respect to the conditioning order of the Rosenblatt transformation is by no means specific to this transformation. There is

no explicit conditioning order to choose in the classical Nataf transformation, but it is in fact hidden in the choice of $\boldsymbol{\Gamma}$ in its definition. The usual choice based on the Cholesky factor $\boldsymbol{L}$ of the correlation matrix $\boldsymbol{R}$ corresponds to the canonical conditioning ordering for the Rosenblatt transformation, but other choices are possible. In fact, the set of possible choices for $\boldsymbol{\Gamma}$ is exactly the set $\{\boldsymbol{Q}\,\boldsymbol{L}^{-1} \,|\, \boldsymbol{Q} \in \mathcal{O}_n(\mathbb{R})\}$. This way, we can recover the Rosenblatt transformation with a non-canonical conditioning ordering and even other transformations. As soon as $\boldsymbol{R} \neq \boldsymbol{I}_n$, each different choice for $\boldsymbol{\Gamma}$ will lead to different importance factors that will not reduce to permutations of the one obtained using the Cholesky factor.

The actual difficulty is in fact not in the choice of the conditioning ordering, but rather to define importance factors that are invariant by permutation in the case of random vector with dependent components. This definition is still an open question to the best of our knowledge.

Let us recall that the exact value of the event probability remains unchanged whatever the transformation we use, and whatever the conditioning order we use for the Rosenblatt transformation!

5.5 Numerical applications

In this section, we illustrate the results obtained in the previous sections through two numerical applications.
We consider a bi-dimensional random vector $\boldsymbol{X} = (X_1, X_2)$ defined by its marginal distribution functions (F_1, F_2) and its copula $C_{\boldsymbol{X}}$.
For both applications, we choose exponential distributions $X_1 \sim \mathcal{E}(\lambda_1)$ and $X_2 \sim \mathcal{E}(\lambda_2)$ for the marginal distributions and a limit state surface defined by:

$$8X_1 + 2X_2 - 1 = 0 \tag{5.34}$$

We consider the event :

$$8X_1 + 2X_2 - 1 \leq 0 \tag{5.35}$$

which we want to evaluate the probability.

In the first application, we choose a normal copula $C_{\boldsymbol{X}} = C^N_{\boldsymbol{R}}$ where $\boldsymbol{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho \in [-1, 1]$ the correlation coefficient of the underlying generic representative of the copula. For $\rho = 0$, the normal copula is the independent copula, for $\rho = -1$ it is the Fréchet lower bound copula and for $\rho = 1$ it is the min copula.

Using the Fréchet-Hoeffding bounds (1.30) and the definition of the linear correlation using copulas (1.14), we can compute the linear correlation ρ_{12} between X_1 and X_2 as a function of ρ. As ρ_{12} is invariant by affine transformations, it does not depend on λ_1 and λ_2, so we take $\lambda_1 = \lambda_2 = 1$ in the computation. We get that ρ_{12} is an increasing function of ρ, with a minimum value of:

$$\rho_{12}^{min} = \int_0^{+\infty} -x \log\left(1 - e^{-x}\right) e^{-x}\, \mathrm{d}x - 1 = \int_0^1 \log(u(1-u))\, \mathrm{d}u - 1 = 1 - \pi^2/6 \simeq -0.645$$

for $\rho = -1$ (X_1 and X_2 are counter-monotone) and a maximum value of

$$\rho_{12}^{max} = \int_0^{+\infty} x^2 e^{-x}\, \mathrm{d}x - 1 = 1$$

for $\rho = 1$ (X_1 and X_2 are co-monotone). This example shows that ρ_{12} cannot be specified independently from the marginal distributions when one uses the normal copula: any value below ρ_{12}^{min} is **not compatible** with the given marginals and the normal copula. The evolution of ρ_{12} as a function of ρ is given on Figure 5.2.

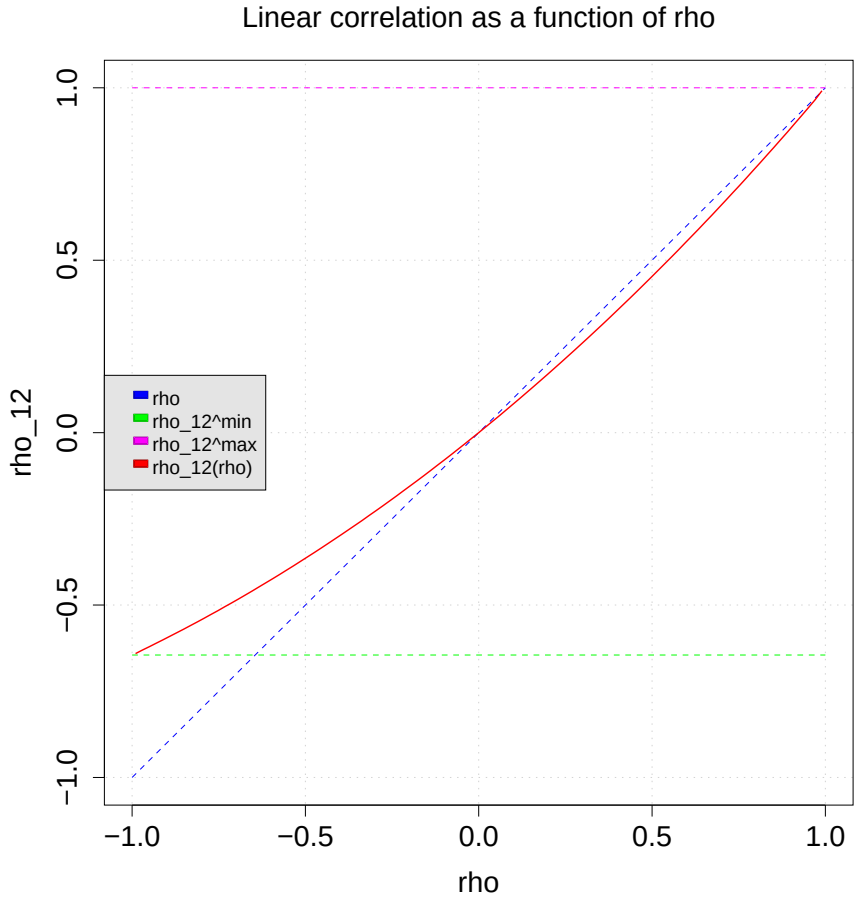


Figure 5.2: Linear correlation $\rho_{12} = (\mathbf{Cor}\,[\boldsymbol{X}])_{12}$ as a function of ρ for exponential marginal distributions $X_1 \sim \mathcal{E}(1)$, $X_2 \sim \mathcal{E}(1)$ and a normal copula with correlation ρ.

In the first application (subsection 5.5.1), we check both the equivalence between the canonical Rosenblatt transformation and the generalized Nataf transformation, and the effect of a change in the conditioning order.

In the second application (subsection 5.5.2), we choose a non-elliptical copula, namely the Frank copula $C_{\boldsymbol{X}} = C_\theta$, which belongs to the class of Archimedean copulas, and we verify that a change in the conditioning order is not equivalent to an orthogonal modification of the transformation and has an impact on the FORM and SORM approximations.

We recall that the Frank copula is defined on $[0,1]^2$ by the expression:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1}{e^{-\theta} - 1}\right) \tag{5.36}$$

where $\theta \in \mathbb{R}^*$. For $\theta \to 0$, the Frank copula tends to the independent copula, for $\theta \to -\infty$

it tends to the Fréchet lower bound copula and for $\theta \to +\infty$ it tends to the min copula.

We can compute the linear correlation ρ_{12} between X_1 and X_2 as a function of θ the same way we did for the preceeding example. We get that ρ_{12} is an increasing function of θ, with the same minimum value of $\rho_{12}^{min} = 1 - \pi^2/6 \simeq -0.645$ for $\theta \to -\infty$ and the same maximum value of $\rho_{12}^{max} = 1$ for $\theta \to +\infty$. The evolution of ρ_{12} as a function of θ is given on Figure 5.3.

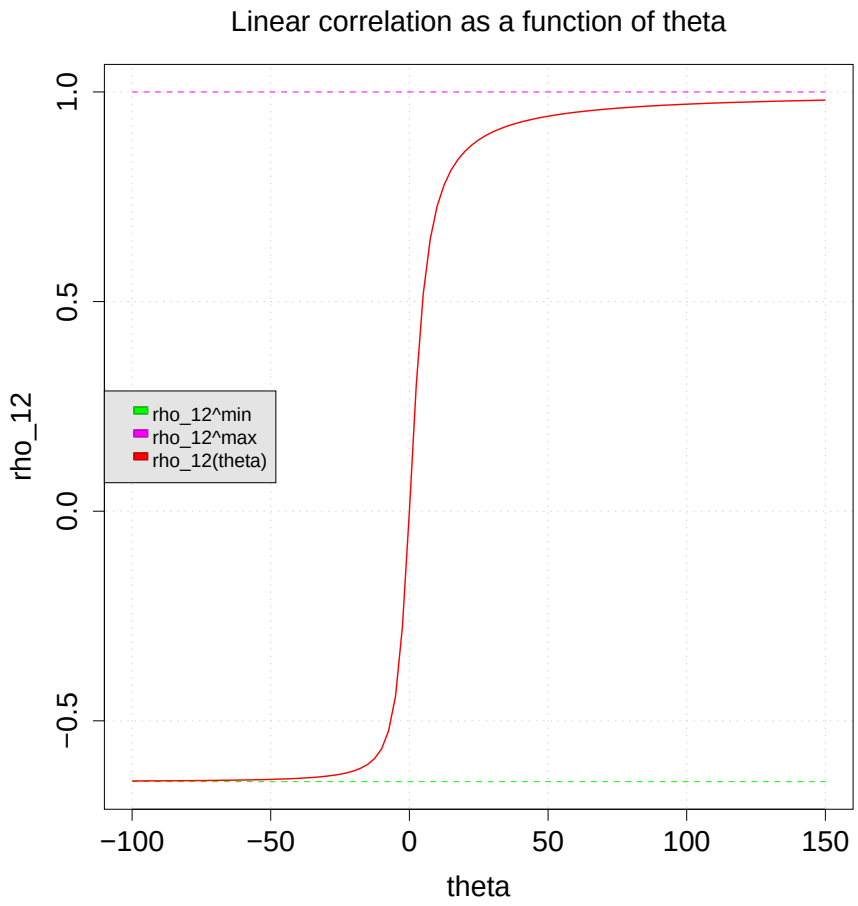


Figure 5.3: Linear correlation $\rho_{12} = (\mathbf{Cor}[\boldsymbol{X}])_{12}$ as a function of θ for exponential marginal distributions $X_1 \sim \mathcal{E}(1)$, $X_2 \sim \mathcal{E}(1)$ and a Frank copula C_θ.

In the numerical applications, we take $\lambda_1 = 1$, $\lambda_2 = 3$, $\rho = 1/2$ and $\theta = 10$.

5.5.1 Application 1: normal copula

We use the new expression of the Rosenblatt transformation of Definition 5.3, with the previous notation $\boldsymbol{W} = T_0(\boldsymbol{X})$, in that particular case of normal copula.
Given Proposition 5.5, the conditional distribution of W_2 given W_1 is a normal distribution such that $\mathbb{E}[W_2|W_1] = \rho W_1$ and $\mathbf{Var}[W_2|W_1] = 1 - \rho^2$, which implies that $F^{W_2|W_1}(W_2|W_1) = \Phi\left(\dfrac{W_2 - \rho W_1}{\sqrt{1-\rho^2}}\right)$.

Finally, the random vector $\boldsymbol{U}$ is defined by:

$$\left\{\begin{array}{rcl} U_1 &=& \Phi^{-1}\circ F^{W_1}(W_1) = W_1 \\ U_2 &=& \Phi^{-1}\circ F^{W_2|W_1}(W_2|W_1) = \dfrac{W_2-\rho W_1}{\sqrt{1-\rho^2}} \end{array}\right. \tag{5.37}$$

The Rosenblatt transformation with canonical order on the conditionning step finally defines the normal random vector $\boldsymbol{U}$ as:

$$\left\{\begin{array}{rcl} U_1 &=& \Phi^{-1}\circ F^{1}(X_1) \\ U_2 &=& \dfrac{\Phi^{-1}\circ F^{2}(X_2) - \rho\Phi^{-1}\circ F^{1}(X_1)}{\sqrt{1-\rho^2}} \end{array}\right. \tag{5.38}$$

In the Rosenblatt standard space, the limit state surface has the parametric expression, where $\xi \in [0,+\infty)$:

$$\left\{\begin{array}{rcl} u_1 &=& \Phi^{-1}\circ F^{1}(\xi) \\ u_2 &=& \dfrac{\Phi^{-1}\circ F^{2}\left(\frac{1-8\xi}{2}\right) - \rho\Phi^{-1}\circ F^{1}(\xi)}{\sqrt{1-\rho^2}} \end{array}\right. \tag{5.39}$$

With the same considerations, the Rosenblatt transformation with the inverse order on the conditioning step defines the normal random vector $\tilde{\boldsymbol{U}}$ as:

$$\left\{\begin{array}{rcl} \tilde{U}_1 &=& \Phi^{-1}\circ F^{2}(X_2) \\ \tilde{U}_2 &=& \dfrac{\Phi^{-1}\circ F^{1}(X_1) - \rho\Phi^{-1}\circ F^{2}(X_2)}{\sqrt{1-\rho^2}} \end{array}\right. \tag{5.40}$$

which leads, in the standard space, to the expression of the limit state surface:

$$\left\{\begin{array}{rcl} \tilde{u}_1 &=& \Phi^{-1}\circ F^{2}(\frac{1-8\xi}{2}) \\ \tilde{u}_2 &=& \dfrac{\Phi^{-1}\circ F^{1}(\xi) - \rho\Phi^{-1}\circ F^{2}\left(\frac{1-8\xi}{2}\right)}{\sqrt{1-\rho^2}} \end{array}\right. \tag{5.41}$$

Figure 5.4 draws the graph of the limit state surface in the standard space after both Rosenblatt transformations.

Thanks to relation (5.25), we can explicit the orthogonal matrix $\boldsymbol{Q}$. The permutation matrix is $\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ which leads to $\boldsymbol{R}_2 = \boldsymbol{R}$. Furthermore, we have $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_2 = \begin{pmatrix} 1 & 0 \\ \dfrac{-\rho}{\sqrt{1-\rho^2}} & \dfrac{1}{\sqrt{1-\rho^2}} \end{pmatrix}$ and finally $\boldsymbol{Q} = \begin{pmatrix} \rho & \sqrt{1-\rho^2} \\ \dfrac{}{\sqrt{1-\rho^2}} & -\rho \end{pmatrix}$.

We can easily verify that $\tilde{\boldsymbol{U}} = \boldsymbol{Q}\boldsymbol{U}$. Furthermore, $\boldsymbol{Q}$ is a permutation matrix whith $\det(\boldsymbol{Q}) = -1$, as the matrix $\boldsymbol{P}$.

The director vector of the symmetry axis is $\left(\sqrt{\dfrac{1+\rho}{2}}, \sqrt{\dfrac{1-\rho}{2}}\right)$. In the numerical application drawn in Figure 5.4, the symmetry axis is $(\sqrt{3}/2, 1/2)$.

The Hasofer reliability index is $\beta = 1.30$ and the FORM approximation of the event probability $p = \mathbb{P}(8X_1 + 2X_2 - 1 < 0)$ is:

$$p_{FORM} = \Phi^{-1}(-\beta) = 9.76\,10^{-2} \tag{5.42}$$

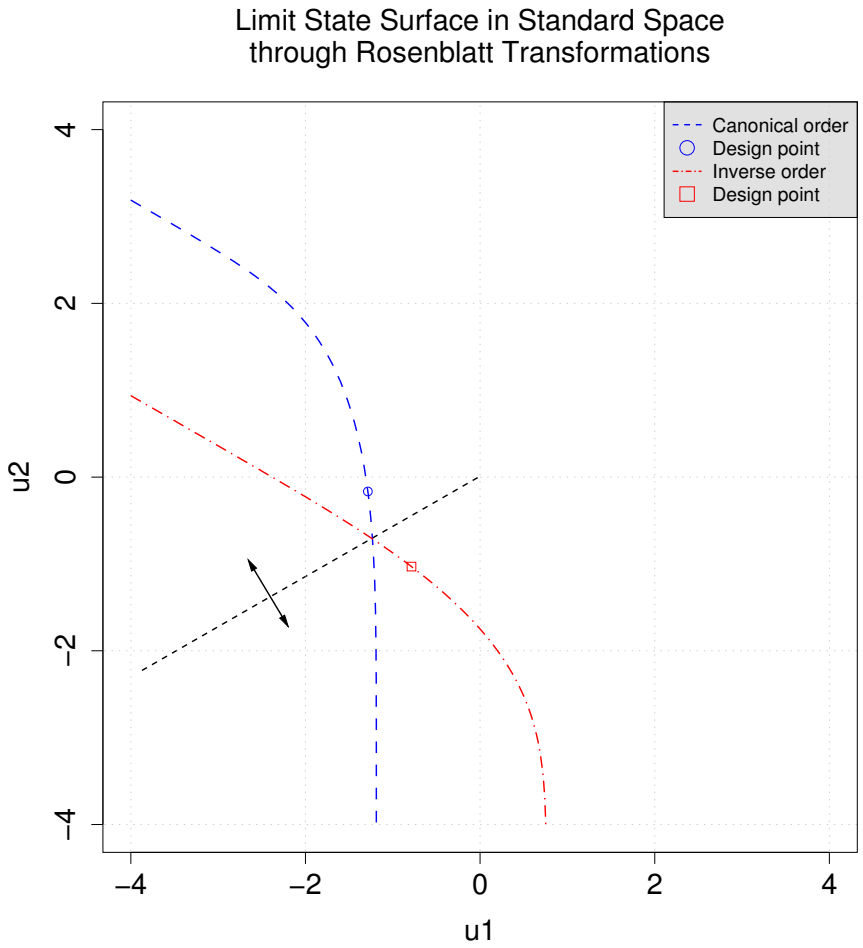


Figure 5.4: Transformations of the limit state surface into the standard space when using the canonical order in the Rosenblatt transformation and its inverse. The linear correlation is $\rho = 1/2$, the copula is normal, $X_1 \sim \mathcal{E}(1)$ and $X_2 \sim \mathcal{E}(3)$. The limit state surface is $8X_1 + 2X_2 - 1 = 0$. Note the symmetry that exchanges the two curves: its matrix is $\boldsymbol{Q}$.

An analytical computation of p leads to the numerical result :

$$p = 8.73\,10^{-2} \tag{5.43}$$

Let us verify now the equivalence between the Nataf transformation and the Rosenblatt one (given that we consider the canonical order). The Nataf transformation leads to the normal random vector $\boldsymbol{U}$ defined as:

$$\boldsymbol{U} = \boldsymbol{\Gamma} \begin{pmatrix} \Phi^{-1} \circ F^1(X_1) \\ \Phi^{-1} \circ F^2(X_2) \end{pmatrix} \tag{5.44}$$

As $\boldsymbol{\Gamma} = \begin{pmatrix} 1 & 0 \\ -\dfrac{\rho}{\sqrt{1-\rho^2}} & \dfrac{1}{\sqrt{1-\rho^2}} \end{pmatrix}$, we have:

$$\begin{cases} U_1 &= \Phi^{-1} \circ F^1(X_1) \\ U_2 &= -\dfrac{\rho \Phi^{-1} \circ F^1(X_1)}{\sqrt{1-\rho^2}} + \dfrac{\Phi^{-1} \circ F^2(X_2)}{\sqrt{1-\rho^2}} \end{cases} \tag{5.45}$$

which is identical to the expression defined in (5.38).

5.5.2 Application 2: Frank copula

We consider here the Frank copula, which is an non-elliptical copula. This example proves that both limit state surfaces in the standard space associated to two different orders in the conditioning step of the Rosenblatt transformation are not linked by an orthogonal transformation. We also illustrate that, according to this conditioning order, the reliability indices are different which leads to different FORM approximations of the probability.

Figure 5.5 draws the graph of the limit state function in the standard space after both Rosenblatt transformations.

The respective reliability index are different in both cases:

$$\begin{cases} \beta_{CanOrd} &= 1.24 \\ \beta_{InvOrd} &= 1.17 \end{cases} \tag{5.46}$$

which leads to different FORM approximations of the event probability:

$$\begin{cases} P_{CanOrd}^{FORM} &= 1.07\,10^{-1} \\ P_{InvOrd}^{FORM} &= 1.22\,10^{-1} \end{cases} \tag{5.47}$$

There is a difference of 14% between the two approximations, only due to the conditioning order, whereas the exact probability value is the same.

An analytical computation of p leads to the numerical value:

$$p = 1.038\,10^{-1} \tag{5.48}$$

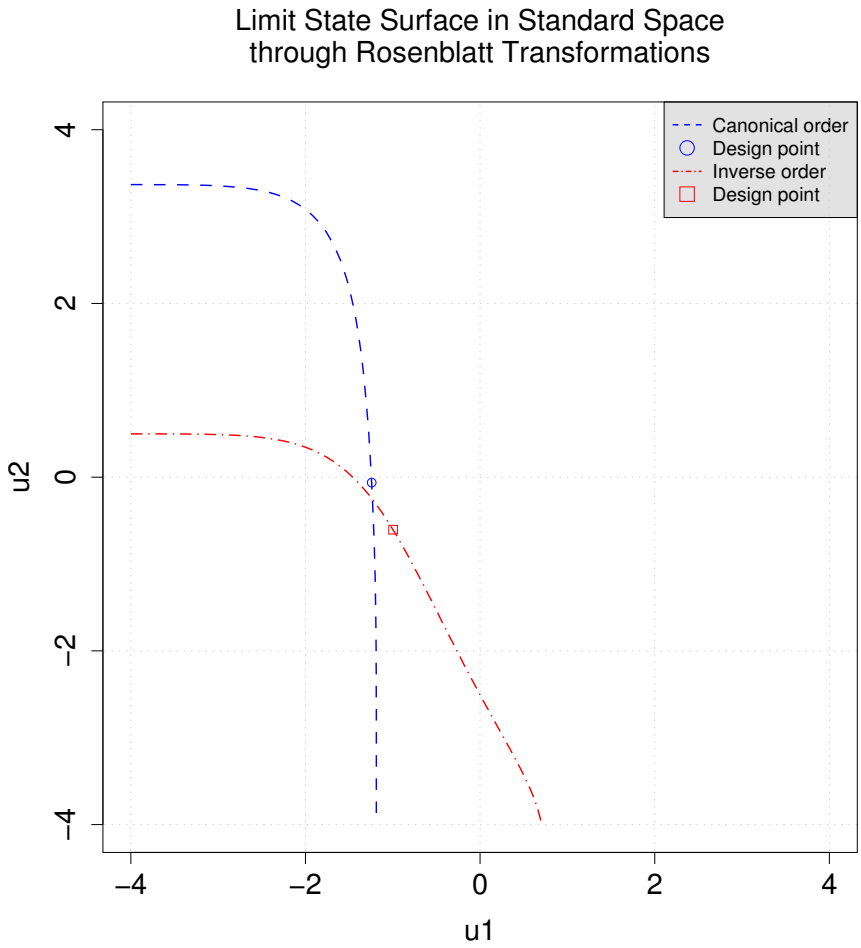


Figure 5.5: Transformations of the limit state surface into the standard space when using the canonical order in the Rosenblatt transformation and its inverse. The copula is a Frank one with $\theta = 10$, $X_1 \sim \mathcal{E}(1)$ and $X_2 \sim \mathcal{E}(3)$. The limit state surface is $8X_1 + 2X_2 - 1 = 0$.

5.6 Conclusion

This chapter concludes our global reflections on isoprobabilistic transformations.

Its first main objective was to compare the generalized Nataf transformation with the Rosenblatt transformation and to show that in the normal copula case, both transformations are identical.

In the use of the Rosenblatt transformation, there is a degree of freedom in the ordering of the conditioning step. This point is often presented as a drawback of this transformation, as it leads to different numerical results for the FORM and SORM approximation. The second main objective of this chapter was to show that, although the conditioning order has such an impact in general, in the normal copula case there is indeed no impact on the FORM and SORM approximations as well as on the reliability index. The only impact is on the importance factors in the case of correlated components for $\boldsymbol{X}$, which underline the difficulty to interpret such factors in the correlated case.

The Nataf transformation has been successfully generalized to produce more general standard spaces than the normal one. We showed that the Rosenblatt transformation cannot be generalized this way. Thus, for the case of a non-normal elliptical copula, one can choose between two isoprobabilistic transformations: the Rosenblatt transformation or the generalized Nataf transformation.

We illustrated these results through two numerical applications, showing the equivalence of both transformations in the normal copula case and the effect of the conditioning order in a normal and non-normal copula case.

Let us recall that the exact value of the event probability remains the same whatever the transformation we use, and whatever the conditioning order we use for the Rosenblatt transformation. It is only the FORM and SORM approximations that are potentially modified.

Chapter 6

Copulas for order statistics with prescribed marginal distribution functions

The probabilistic modeling of a random vector is not always based on information about the marginal distributions and the dependence structure. In some cases, in addition to marginal information, we have an information about the support of the joint distribution. This additional information usually constrains both the marginal distributions and the copula. In this chapter, we study in details this kind of modeling situation in the case of constraints that reduce to a non-decreasing ordering that must be satistfied almost surely.

The main results of this chapter are the characterization in the absolutely continuous case of the one-dimensional marginal distributions and the n-dimensional copulas compatible with such constraints, and the definition of a new infinite-dimensional parametric family of copulas well-suited to this modeling situation. We prove the existence and the uniqueness of a copula with maximal support within this family in the bi-dimensional case, and give efficient algorithms to work with this copula.

6.1 Introduction

Modeling the joint distribution of a random vector with prescribed marginal distributions and partial dependence information was a challenging task before the wide diffusion of the copula concept. Thanks to the Sklar Theorem 1.11, we know that it is a problem of copula selection: any copula able to reproduce this partial dependence information will lead to a plausible multivariate distribution function. When the partial information is given through a set of scalar measures of association, many procedures are available to select a parametric family of copulas and to estimate its parameters using either estimates of measures of association as presented in chapter 1, or more specific methods as presented in [Nel06], [Joe97]. When one has a multivariate sample at hand, it is even possible to resort to semi-parametric or non-parametric estimations of the copula [CFS07], [GN07].

But in some modeling situations, the dependence information is not given explicitely but through a constraint that must be fulfilled almost surely by the random vector. More precisely, if $B \subset \mathbb{R}^n$ is a given measurable set, we are interested in the characterization of n-dimensional distribution functions G with prescribed marginal distribution functions

$G_1, \dots, G_n$ for $\boldsymbol{X}$ such that:

$$\boldsymbol{X} \in B \text{ a.s.} \tag{6.1}$$

This problem is known as the marginals problem (see e.g. [RT94] and [Str65]) and has been studied in terms of variational conditions, but does not seem to have been extensively studied from the viewpoint of copulas in the literature. We propose to adress it in the particular setting where all the marginal distribution functions $G_1, \dots, G_n$ are continuous. We also restrict our attention to domains B such that:

$$B = \{\boldsymbol{x} \in \mathbb{R}^n \,|\, \psi_1(x_1) \leq \dots \leq \psi_n(x_n)\}$$

where $(\psi_i)_{i=1,\dots,n}$ are strictly monotonic continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

This problem can be reduced to the case where $\forall i = 1, \dots, n, \forall x_i \in \mathbb{R}, \quad \psi_i(x_i) = x_i$: if we define $F_i(x_i) = G_i(\psi_i^{-1}(x_i))$ if ψ_i is strictly increasing and $F_i(x_i) = 1 - G_i(\psi_i^{-1}(x_i))$ if ψ_i is strictly decreasing, then there is a bijection between the joint distribution functions F of random vectors $\boldsymbol{X}$ with continuous marginal distribution functions $F_1, \dots, F_n$ and satisfying:

$$X_1 \leq \dots \leq X_n \text{ a.s} \tag{6.2}$$

and the joint distribution functions of the random vectors $\boldsymbol{X}$ with continuous marginal distribution functions $G_1, \dots, G_n$ and satisfying (6.1).

This new formulation shows that the problem reduces to the determination of the admissible copulas for the joint distribution function F of order statistics with prescribed marginal distributions $F_1, \dots, F_n$. The large literature on order statistics is mainly focused on the derivation of F given an initial random vector $\boldsymbol{X}$ with unordered components and a given joint distribution function H, the most common setting being random vectors with identical marginal distribution functions $H_1, \dots, H_n = H_{ref}$ and independent or exchangeable components (see [ABNN08], [DN03]), the most involved setting being the case of arbitrary marginal distribution functions $H_1, \dots, H_n$ and independent components wich leads to the Bapat-Beg theorem [BB89] which fully describes F in terms of the H_i. More recently, some authors started to link order statistics and copulas such as in [JR08], and to study the dependence structure of order statistics [AGK05], but with a focus on measures of concordance rather than on admissible copulas. The closest published work is [NS10], where the authors characterize the copulas associated with H when the marginal distribution functions $F_1, \dots, F_n$ are prescribed, whereas we are interested in the admissible copulas associated with F for given $F_1, \dots, F_n$.

The chapter is organized as follows. We start by recalling the well known compatibility conditions for continuous marginal distribution functions to be the marginal distribution functions of order statistics and we give a synthetic table showing the pairs of marginal distribution functions which do not satisfy this compatibility condition regardless of their parameterization. Then, we fully characterize the set of copulas compatible with (6.2) when the marginal distribution functions are also compatible. This characterization excludes all the classical absolutely continuous copulas, so we define a new family of copulas dedicated to this modeling situation. We prove the existence and uniqueness of the copula of largest support in this family, and we give all the algorithmic details that allows to sample such a copula, e.g. in order to perform Monte Carlo simulations. In the conclusion, we mention some possible extensions to this work.

6.2 Compatibility conditions

The joint distribution of order statistics has been a long term research topic, but mainly focused on the derivation of this distribution given the distribution of an initial unordered

random vector. It is only recently that some work has been published concerning the possible structure of the joint distribution of order statistics with no reference to a parent distribution. While the necessary condition on its marginal distribution functions is a well-known result, the fact that it is also a sufficient condition is not so well known, and the characterization of compatible copulas has not been described so far.

In the next paragraph we recall the compatibility condition on the marginal distributions and give a synthetic view of it for classical marginal distributions in Table 6.1.

6.2.1 Conditions on marginals

In order to study the general case where the different continuous marginal distribution functions are increasing but not strictly increasing, we have to define two generalized inverse for such functions:

Definition 6.1. Let F be an increasing function from $\mathbb{R}$ to $[a,b]$. We define the two generalized inverse $F^{\leftarrow}$ and $F^{\rightarrow}$ by:

$$\begin{aligned} \forall q \in [a,b], \quad F^{\leftarrow}(q) &= \inf\{x \in \overline{\mathbb{R}} \,|\, F(x) \geq q\} \\ F^{\rightarrow}(q) &= \sup\{x \in \overline{\mathbb{R}} \,|\, F(x) \leq q\} \end{aligned}$$

These two generalized inverse functions are increasing from $[a,b]$ to $\overline{\mathbb{R}}$, and we have:

$$\begin{aligned} \forall x \in \mathbb{R}, \quad F^{\leftarrow}(F(x)) &\leq x \\ F^{\rightarrow}(F(x)) &\geq x \end{aligned}$$

and

$$\begin{aligned} \forall q \in [a,b], \quad F(F^{\leftarrow}(q)) &\geq q \\ F(F^{\rightarrow}(q)) &\leq q \end{aligned}$$

these last two inequalities being equalities if F is continuous.

When applied to a distribution function F, we have $a = 0$, $b = 1$ and $F^{\leftarrow}$ is equal to the quantile function F^{-1} associated with F, see Definition 1.5.

We recall a well-known result on the marginal distribution functions of order statistics, telling that order statistics are also ordered according to the usual stochastic ordering:

Theorem 6.2. *Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random vector satisfying (6.2). Then its marginal distribution functions $(F_1, \ldots, F_n)$ verify the point-wise inequality:*

$$\forall i \in \{1, \ldots, n-1\},\ \forall x \in \mathbb{R},\ F_i(x) \geq F_{i+1}(x) \tag{6.3}$$

or equivalently:

$$\forall i \in \{1, \ldots, n-1\},\ \forall t \in [0,1],\ F_i^{\leftarrow}(t) \leq F_{i+1}^{\leftarrow}(t) \tag{6.4}$$

The condition (6.3) is also known as the **first order dominance** as in [Lev06] or also the **stochastic ordering** of $X_1, \ldots, X_n$ as in [Nel06].

Proof. By (6.2), $\forall i \in \{1, \ldots, n-1\}$ and $\forall x \in \mathbb{R}$, $A_{i+1} = \{X_{i+1} \leq x\} \subset A_i = \{X_i \leq x\}$ from which $F_i(x) = \mathbb{P}(A_i) \geq F_{i+1}(x) = \mathbb{P}(A_{i+1})$.
The condition on the quantile functions is easily obtained from the definition of quasi-inverse and the right-continuity of distribution functions. More precisely, for $t \in [0,1]$, for all $i \in \{1, \ldots, n\}$ we define $A_i(t) = \{x \in \mathbb{R} \,|\, F_i(x) \geq t\}$. We deduce from $F_i \geq F_{i+1}$

that $A_{i+1}(t) \subset A_i(t)$ from which $F_i^{\leftarrow} \leq F_{i+1}^{\leftarrow}$. Conversely, if there exists $\tilde{x} \in \mathbb{R}$ such that $F_i(x) < F_{i+1}(\tilde{x}) = t$, then from the right-continuity of F_i there exists $\epsilon > 0$ such that $\forall x \in [\tilde{x}, \tilde{x}+\epsilon]$, $F_i(x) < t$ and as F_i is non-decreasing, $\inf A_i(t) > \tilde{x} + \epsilon$. Moreover, as F_{i+1} is non-decreasing and $F_{i+1}(\tilde{x}) = t$, $\inf A_i(t) \leq \tilde{x}$, from which $F_i^{\leftarrow}(t) > F_{i+1}^{\leftarrow}(t)$. □

We can deduce from this proposition some easy-to-check necessary compatibility or incompatibility conditions. We define the **bounds** of a random variable as:

Definition 6.3. Let X be a real valued random variable with marginal distribution function F. We define its **lower bound** $\underline{X}$ and its **upper bound** $\overline{X}$ by

$$\underline{X} = F^{\rightarrow}(0) = \sup\{x \in \overline{\mathbb{R}} \,|\, F(x) = 0\}$$
$$\overline{X} = F^{\leftarrow}(1) = \inf\{x \in \overline{\mathbb{R}} \,|\, F(x) = 1\} = F^{-1}(1)$$

Then, a necessary compatibility condition is given by Corollary 6.4.

Corollary 6.4. *If $\boldsymbol{X}$ satisfies the condition (6.2), the bounds of its components must satisfy:*

$$\underline{X}_1 \leq \ldots \leq \underline{X}_n \text{ and } \overline{X}_1 \leq \ldots \leq \overline{X}_n \tag{6.5}$$

Proof. If (6.5) is not fulfilled, there exists a $i \in \{1, \ldots, n-1\}$ such that for example $\underline{X}_{i+1} < \underline{X}_i$. Let x be in $(\underline{X}_{i+1}, \underline{X}_i) \neq \emptyset$. Then we have $F_{X_i}(x) = 0$ and $F_{X_{i+1}}(x) > 0$ which is incompatible with (6.3) and then (6.2). The case $\overline{X}_{i+1} < \overline{X}_i$ is dealt with in the same way. □

By studying the tails of the marginal distribution functions, one can also get sufficient incompatibility conditions as given in Corollary 6.5.

Corollary 6.5. *If $F_i = o(F_{i+1})$ in the vicinity of $\underline{X}_i$ or $1 - F_{i+1} = o(1 - F_i)$ in the vicinity of $\overline{X}_{i+1}$ then the marginal distribution functions F_i and F_{i+1} are not compatible with the constraint (6.2).*

In the case where F_i and F_{i+1} are absolutely continuous, these criteria can be translated in terms of marginal density functions p_i and p_{i+1}: if $p_i = o(p_{i+1})$ in the vicinity of $\underline{X}_i$ or $p_{i+1} = o(p_i)$ in the vicinity of $\overline{X}_{i+1}$ then the marginal distribution functions F_i and F_{i+1} are not compatible with the constraint (6.2).

Proof. First, we note that $F_i = o(F_{i+1})$ implies that $\underline{X}_{i+1} \leq \underline{X}_i$, else we would have $F_i > 0$ and $F_{i+1} \equiv 0$ on $[\underline{X}_i, \underline{X}_{i+1})$, which is not compatible with $F_i = o(F_{i+1})$.

In the vicinity of $\underline{X}_i$, the condition $F_i = o(F_{i+1})$ translates into $\forall \epsilon > 0$ and $\epsilon < 1$, $\exists \eta > 0$ such that $\forall x \in [\underline{X}_i, \underline{X}_i + \eta)$, $F_i(x) \leq \epsilon F_{i+1}(x) < F_{i+1}(x)$ which is incompatible with (6.3). The symmetric condition in the vicinity of $\overline{X}_{i+1}$ is dealt with the same way. The condition expressed in terms of marginal densities reduces to the condition on one-dimensional marginal distribution functions by integration on $(-\infty, x]$. □

If we consider the set of usual distributions given in Table 6.1, then using Corollaries 6.4 and 6.5, we can discard half of the couples of distributions as being incompatible with the constraint, regardless of the set of parameters we choose for these distributions. For the remaining couples, it may be possible to impose the constraint but with restrictions on the values of the parameters. For example, in the case of two normal distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, we must have $\mu_1 \leq \mu_2$ and $\sigma_1^2 = \sigma_2^2$, whereas for two uniform distributions $\mathcal{U}(a_1, b_1)$ and $\mathcal{U}(a_2, b_2)$, we must have $a_1 \leq a_2$ and $b_1 \leq b_2$.

We conclude that the constraint (6.2) imposes strong conditions both on the family of marginal distributions and on the possible values for their parameters. These conditions have to be taken into account when one tries to infer marginal distributions from statistical data if the underlying multivariate quantity is known to satisfy such a constraint.

	Beta	Exponential	Gamma	Gumbel	Laplace	Logistic	LogNormal	Normal	Rayleigh	Student	Triangular	Uniform	Weibull
Beta				∅	∅			∅		∅			
Exponential	∅			∅	∅			∅	∅	∅	∅	∅	
Gamma	∅			∅	∅			∅	∅	∅	∅	∅	
Gumbel	∅							∅	∅	∅	∅	∅	
Laplace	∅							∅	∅		∅	∅	
Logistic	∅			∅	∅			∅	∅	∅	∅	∅	
LogNormal	∅			∅	∅	∅		∅	∅	∅	∅	∅	∅
Normal	∅			∅	∅					∅	∅	∅	
Rayleigh	∅			∅	∅			∅		∅	∅	∅	
Student	∅	∅	∅	∅	∅	∅	∅	∅	∅		∅	∅	∅
Triangular				∅	∅			∅		∅			
Uniform				∅	∅			∅		∅			
Weibull										∅	∅	∅	

Table 6.1: Possible couples of marginal distributions for (X_i, X_{i+1}) where the distribution of X_i is given by rows and those of X_{i+1} is given by columns. The ∅ symbol means the marginal distributions are not compatible with the constraint (6.2), regardless of their parameters. In the other cases, there exists a possible combination of parameters compatible with the constraint (6.2). For example, one cannot have an exponential distribution for X_i and a uniform distribution for X_{i+1}, but the converse may be possible for specific values of the parameters of these distributions.

6.2.2 Conditions on copula

The main result of this section is Theorem 6.7 which characterizes the set of copulas of order statistics with prescribed compatible marginal distributions. We will denote by μ_F the unique positive measure defined on the σ-field $\mathcal{B}(\mathbb{R}^n)$ associated with the distribution function F, and by μ_C the unique positive measure defined on the σ-field $\mathcal{B}([0,1]^n)$ associated with the copula C.

For the remainder of the chapter, $F_1, \ldots, F_n$ are n marginal distribution functions verifying (6.3). Let $\mathcal{C}$ be the set of n-dimensional multivariate copulas and $\mathcal{C}(F_1, \ldots, F_n)$ the set of multivariate copulas which, associated with $F_1, \ldots, F_n$ lead to multivariate distribution functions verifying (6.2). The next result shows that the copula selection problem is well-posed, in the sense that if (6.3) is satisfied by $F_1, \ldots, F_n$, then there exist copulas compatible with $F_1, \ldots, F_n$, i.e $\mathcal{C}(F_1, \ldots, F_n) \neq \emptyset$.

Proposition 6.6. *Let M_n be the* **min copula** *of dimension n, defined by*

$$\forall \boldsymbol{u} \in [0,1]^n,\ M_n(\boldsymbol{u}) = \min(u_1, \ldots, u_n)$$

then $M_n \in \mathcal{C}(F_1, \ldots, F_n)$ so it is always possible to build a joint distribution function of order statistics with compatible marginal distribution functions.

Proof. Let U be a random variable uniformly distributed on $[0,1]$ and $\boldsymbol{X} = (F_1^{\leftarrow}(U), \ldots, F_n^{\leftarrow}(U))$. By construction, M_n is an admissible copula for $\boldsymbol{X}$ and its marginal distribution functions are $F_1, \ldots, F_n$. By construction, using (6.4), $\boldsymbol{X}$ satisfies the constraint (6.2). □

We introduce three sets that will play a key role in the study of $\mathcal{C}(F_1, \ldots, F_n)$:

$$\mathcal{S}_X = \left\{ \boldsymbol{x} \in \overline{\mathbb{R}}^n \,|\, x_1 \leq \ldots \leq x_n \right\} = \bigcap_{i=1}^{n-1} \left\{ \boldsymbol{x} \in \overline{\mathbb{R}}^n \,|\, x_i \leq x_{i+1} \right\} \tag{6.6}$$

$$\mathcal{S}_U = \left\{ \boldsymbol{u} \in [0,1]^n \,|\, u_1 \leq \ldots \leq u_n \right\} = \bigcap_{i=1}^{n-1} \left\{ \boldsymbol{u} \in [0,1]^n \,|\, u_i \leq u_{i+1} \right\} \tag{6.7}$$

$$\Delta(F_1, \ldots, F_n) = T(\mathcal{S}_X) \subset [0,1]^n \tag{6.8}$$

where T is the usual probability integral transformation (see e.g [Rüs09]), defined by:

$$T : \left| \begin{array}{ccc} \mathbb{R}^n & \to & [0,1]^n \\ \boldsymbol{x} & \mapsto & \boldsymbol{u} = \begin{pmatrix} F_1(x_1) \\ \vdots \\ F_n(x_n) \end{pmatrix} \end{array} \right.$$

We give now the characterization of the copulas of order statistics given compatible continuous marginal distributions.

Theorem 6.7. *The set of copulas $\mathcal{C}(F_1, \ldots, F_n)$ is exactly the set of copulas C such that:*

$$\mu_C(\Delta(F_1, \ldots, F_n)) = 1$$

it means, copulas whose support is included in $\Delta(F_1, \ldots, F_n)$.

Proof. Let F be a multivariate distribution function with marginal distribution functions $F_1, \ldots, F_n$ and copula C. C is in $\mathcal{C}(F_1, \ldots, F_n)$ if and only if F satisfies the constraint 6.2, which means $\mu_F(\mathcal{S}_X) = 1$. By Theorem 1.11, we have:

$$\mu_F(\mathcal{S}_X) = \mu_{C \circ T}(\mathcal{S}_X) = \mu_C(\Delta(F_1, \ldots, F_n))$$

so $\mu_F(\mathcal{S}_X) = 1$ if and only if $\mu_C(\Delta(F_1, \ldots, F_n)) = 1$. □

An immediate corollary of Theorem 6.7 is the following:

Corollary 6.8. $\Delta(F_1, \ldots, F_n) = [0,1]^n$, *i.e. any copula is admissible, if and only if:*

$$\forall i \in \{1, \ldots, n\},\ \overline{X}_i \leq \underline{X}_{i+1} \tag{6.9}$$

Proof. It is obvious that if we have (6.9), then the independent copula Π_n of dimension n is in $\Delta(F_1, \ldots, F_n)$. As its support is $[0,1]^n$, it shows that $\Delta(F_1, \ldots, F_n) = [0,1]^n$.

On the other hand, if $\Delta(F_1, \ldots, F_n) = [0,1]^n$, the independent copula Π is in $\mathcal{C}(F_1, \ldots, F_n)$. Let $\boldsymbol{X}$ be a random vector with marginal distribution functions $F_1, \ldots, F_n$ and copula Π. If there exists $i \in \{1, \ldots, n\}$ such that $\underline{X}_{i+1} < \overline{X}_i$, then for any $s \in \mathbb{R}$ we have:

$$\mathbb{P}(X_i > X_{i+1}) \geq \mathbb{P}(X_i > s \cap X_{i+1} \leq s) = (1 - F_i(s))F_{i+1}(s)$$

If we take $s = \frac{\underline{X}_{i+1} + \overline{X}_i}{2}$, we get $\mathbb{P}(X_i > X_{i+1}) > 0$, which contradicts the constraint. □

As the compatibility condition is expressed as a constraint on the support of the copula and as the set of n-dimensional copulas is convex, one gets the following corollary which allows to construct new copulas in $\mathcal{C}(F_1, \ldots, F_n)$ from given ones:

Corollary 6.9. $\mathcal{C}(F_1, \ldots, F_n)$ *is a convex subset of* $\mathcal{C}$.

The geometry of $\Delta(F_1, \ldots, F_n)$ plays a key role in the study of $\mathcal{C}(F_1, \ldots, F_n)$. The next result shows that for a given n, $\Delta(F_1, \ldots, F_n)$ cannot be arbitrary small as it contains a subset independent of $F_1, \ldots, F_n$ with positive Lebesgue measure.

Proposition 6.10. *We have:*

$$\mathcal{S}_U \subset \Delta(F_1, \ldots, F_n)$$

Proof. For all $\boldsymbol{u} \in \mathcal{S}_U$, if we take $\boldsymbol{x} = (F_1^{\leftarrow}(u_1), \ldots, F_n^{\leftarrow}(u_n))$, from the definition of $\mathcal{S}_U$ and the property (6.4) we have $\boldsymbol{x} \in \mathcal{S}_X$. The continuity of the distribution functions F_i gives $T(\boldsymbol{x}) = \boldsymbol{u}$, so $\boldsymbol{u} \in \Delta(F_1, \ldots, F_n)$. □

The next proposition characterizes the situations where $\Delta(F_1, \ldots, F_n)$ is the smallest possible set:

Proposition 6.11. *The following properties are equivalent:*

1. $\mathcal{C}(F_1, \ldots, F_n) = \{M_n\}$
2. $\forall x \in \mathbb{R}, F_1(x) = \ldots = F_n(x)$
3. $\Delta(F_1, \ldots, F_n) = \mathcal{S}_U$

Proof. We will show that 1⇒ 2⇒ 3⇒ 1.

To show that 1⇒ 2, let us suppose that 2 is not verified. Then there exist $i \in \{1, \ldots, n-1\}$ and $\xi \in \mathbb{R}$ such that $\beta = F_i(\xi) > F_{i+1}(\xi) = \alpha$. We define the set Q by:

$$Q = \{\boldsymbol{u} \in [0,1]^n \,|\, (u_i, u_{i+1}) \in [\alpha, \beta]^2\}$$

and the copula C by:

$$\forall \boldsymbol{u} \in [0,1]^n, C(\boldsymbol{u}) = \begin{cases} \alpha + (\beta - \alpha)C^*\left(\frac{u_1 - \alpha}{\beta - \alpha}, \ldots, \frac{u_n - \alpha}{\beta - \alpha}\right) & \text{if } \boldsymbol{u} \in Q \\ M_n(\boldsymbol{u}) & \text{otherwise} \end{cases}$$

where C^* is the copula defined by:

$$\forall \boldsymbol{u} \in [0,1]^n, C^*(\boldsymbol{u}) = \min_{j \leq i}(u_j) \min_{k \geq i+1}(u_k)$$

C^* is the copula of a random vector for which the i first components are almost surely a strictly increasing function of a random variable U uniformly distributed on $[0,1]$ and its $(n-i)$ remaining components a strictly increasing function of another random variable V uniformly distributed on $[0,1]$, U and V being independent. C is the ordinal sum of C^* with respect to $[\alpha, \beta]$, see [Nel06, Section 3.2.2].

The support Ω of C is:

$$\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$$

with:

$$\begin{aligned}
\Omega_1 &= \{(s, \ldots, s) \in [0,1]^n \,|\, s \in [0, \alpha]\} \\
\Omega_2 &= \{(t, \ldots, t) \in [0,1]^n \,|\, t \in [\beta, 1]\} \\
\Omega_3 &= \{\boldsymbol{u} \in [0,1]^n \,|\, (s,t) \in [\alpha, \beta]^2, \forall j \leq i, u_j = s, \forall k \geq i+1, u_k = t\}
\end{aligned}$$

It remains to show that $\Omega \subset \Delta(F_1, \ldots, F_n)$. If $\boldsymbol{u} \in \Omega_1 \cup \Omega_2$, we take $\boldsymbol{x} = (F_1^{\leftarrow}(u_1), \ldots, F_n^{\leftarrow}(u_n)) \in \overline{\mathbb{R}}^n$ and we have $T(\boldsymbol{x}) = \boldsymbol{u}$ by construction, and $\boldsymbol{x} \in \mathcal{S}_X$ by (6.4) so $\boldsymbol{u} \in \Delta(F_1, \ldots, F_n)$. If $\boldsymbol{u} \in \Omega_3$, we take $\boldsymbol{x} \in \overline{\mathbb{R}}^n$ defined by $\forall j \leq i, x_j = F_j^{\leftarrow}(u_j)$ and $\forall k \geq i+1, x_k = F_k^{\rightarrow}(u_k)$. By construction, $T(\boldsymbol{x}) = \boldsymbol{u}$, and $x_1 \leq \ldots \leq x_i$, $x_{i+1} \leq \ldots \leq x_n$. As $(u_i, u_{i+1}) \in [\alpha, \beta]^2$, $x_i \in [F_i^{\leftarrow}(\alpha), F_i^{\leftarrow}(\beta)]$ and $x_{i+1} \in [F_{i+1}^{\leftarrow}(\alpha), F_{i+1}^{\leftarrow}(\beta)]$ and $F_i^{\leftarrow}(\beta) \leq \xi \leq F_{i+1}^{\leftarrow}(\alpha)$ so $\boldsymbol{x} \in \mathcal{S}_X$ and $\boldsymbol{u} \in \Delta(F_1, \ldots, F_n)$.

The implication 2$\Rightarrow$ 3 is an immediate consequence of the definition of $\Delta(F_1, \ldots, F_n)$ and the fact that the transformation T is the identity transformation thanks to the hypothesis made on $F_1, \ldots, F_n$.

The implication 3$\Rightarrow$ 1 results from the characterization of M_n by its diagonal section. We know from Proposition 6.6 that $M_n \in \mathcal{C}(F_1, \ldots, F_n) \neq \emptyset$. Let C be a copula in $\mathcal{C}(F_1, \ldots, F_n)$.

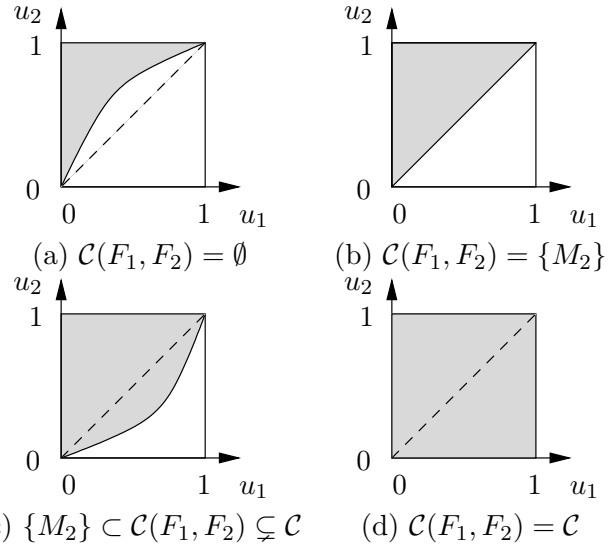
For all $u \in [0,1]$, we have $C(1, \ldots, 1, u) = \mu_C(\{\boldsymbol{v} \in [0,1]^n \,|\, v_n \leq u\}) = u$ by the definition of a copula. We have the partition $\{\boldsymbol{v} \in [0,1]^n \,|\, v_n \leq u\} = \{\boldsymbol{v} \in [0,1]^n \,|\, v_1 \leq u, \ldots, v_n \leq u\} \cup \{\boldsymbol{v} \in [0,1]^n \,|\, \exists i \in \{1, \ldots, n-1\}, v_i > u \text{ and } v_n \leq u\}$. The measure of the first set is equal to $C(u, \ldots, u)$, and the measure of the second set is zero as it corresponds to points $\boldsymbol{v} \in [0,1]^n$ such that $v_i > v_n$, i.e. points in $\mathcal{S}_U^c$. We conclude that $\forall u \in [0,1]$, $C(u, \ldots, u) = u$.

This last property characterizes the copula M_n. Let us take $\boldsymbol{u} \in [0,1]^n$ and define $\tilde{u} = \min\{u_1, \ldots, u_n\} = M_n(\boldsymbol{u})$. We have $\tilde{A} = \{\boldsymbol{v} \in [0,1]^n \,|\, v_1 \leq \tilde{u}, \ldots, v_n \leq \tilde{u}\} \subset A = \{\boldsymbol{v} \in [0,1]^n \,|\, v_1 \leq u_1, \ldots, v_n \leq u_n\}$ so $\mu_C(\tilde{A}) = C(\tilde{u}, \ldots, \tilde{u}) = \tilde{u} = M_n(\boldsymbol{u}) \leq \mu_C(A) = C(\boldsymbol{u})$. On the other hand, Theorem 1.15 give $C(\boldsymbol{u}) \leq M_n(\boldsymbol{u})$, so $C = M_n$. □

Corollary 6.12. *If $\boldsymbol{X}$ is a random vector verifying the constraint (6.2) and with continuous marginal distribution functions $F_1, \ldots, F_n$ all equal to F_0, then we have $X_1 = \ldots = X_n$ a.s.*

Proof. From the definition of $\Delta_{F_0} = \Delta(F_0, \ldots, F_0)$, we have $\Delta_{F_0} \subset \mathcal{S}_U$ and by Proposition 6.10, we have $\Delta_{F_0} = \mathcal{S}_U$ and then the copula of $\boldsymbol{X}$ is M_n. From Theorem 1.16, there exist increasing functions $\psi_2, \ldots, \psi_n$ such that $\forall i \in \{2, \ldots, n\}$, $X_i = \psi_i(X_1)$. We deduce that $\forall i \in \{2, \ldots, n\}$, $\forall x \in \Delta_{F_0}$, $F_0(x) = F_0(\psi_i(x))$. As $\psi_i(\Delta_{F_0}) = \Delta_{F_0}$ and F_0 is invertible on Δ_{F_0}, we have $\forall x \in \Delta_{F_0}, \psi_i(x) = x$, i.e. $\forall i \in \{2, \ldots, n\}$, $X_i = X_1$ *a.s.* □

We illustrate the possible shapes for $\Delta(F_1, F_2)$ and the associated set $\mathcal{C}(F_1, F_2)$ on Figure 6.1. The case (a) corresponds to marginal distributions (F_1, F_2) not compatible with the constraint, as a result of Theorem 6.2. The case (b) corresponds to $F_1 = F_2$ and $C(F_1, F_2) = M_2$ as a result of Proposition 6.11. The case (c) corresponds to $\{M_2\} \subset \mathcal{C}(F_1, F_2) \subset \mathcal{C}$. The case (d) corresponds to $\mathcal{C}(F_1, F_2) = \mathcal{C}$ as a result of Corollary 6.8.



(a) $\mathcal{C}(F_1, F_2) = \emptyset$ (b) $\mathcal{C}(F_1, F_2) = \{M_2\}$

(c) $\{M_2\} \subset \mathcal{C}(F_1, F_2) \subsetneq \mathcal{C}$ (d) $\mathcal{C}(F_1, F_2) = \mathcal{C}$

Figure 6.1: Sketch of the possible shapes for $\Delta(F_1, F_2)$ (grey region) according to the different possible choices for F_1 and F_2. The corresponding set of compatible copulas $\mathcal{C}(F_1, F_2)$ can be empty in case (a), reduced to a single copula in case (b), be a proper subset of all the copulas in case (c) or be the set of all the bi-dimensional copulas in case (d).

We note that the case (c) is the only one where $\mathcal{C}(F_1, F_2)$ is a non-trivial subset of $\mathcal{C}$ and which excludes all the classical continuous copulas as their support is $[0, 1]^2$. In the literature (see [Nel06, Chapter 3]), we find some examples of copulas with restricted supports but none of these copulas are continuous. More precisely, we can find techniques to construct:

- Singular copulas with prescribed support,
- Copulas with limited support from a countable collection of copulas (the **ordinal sum** mechanics),
- Copulas as a convex sum of finite or infinite collection of copulas.

In the first case, we only get singular copulas which are not well suited to screening applications. In the second one, we are limited on the shape of the possible support of the copula: by construction, the support is included inside a union of squares which diagonal is part of the diagonal of $[0, 1]^n$ and share at most one corner. The third case requires having a finite or infinite collection of copulas at hand. By Corollary 6.9, these techniques are of all interest.

The next sections are dedicated to the presentation of a new family of continuous copulas, the **sub-hypercube copulas**, with restricted support.

6.3 Construction of compatible copulas: the sub-hypercube copula family

In this part, we describe a general construction that allows us to build copulas with restricted support from copulas with full support. We explore in more detail the copula obtained starting from the independent copula, and in the bi-dimensional case, we give a theorem on the existence and uniqueness of a copula with the largest possible support within this family.

6.3.1 From copulas with full support to copulas with restricted support

The informal way to construct a copula with restricted support is to chain the following steps:

1. Take a continuous copula H with support $[0,1]^n$;
2. Restrict the copula to a subset $\Delta \subset [0,1]^n$ to get a multivariate distribution function D_Δ;
3. Extract the copula C_Δ of D_Δ: it will have a support that is a subset of $[0,1]^n$, in general a strict subset if Δ is not the cartesian product of two subsets of $[0,1]$.

In this construction, both H and Δ are free parameters. The second step transforms the copula H into a distribution supported by Δ whereas the third step transforms this distribution into a copula with support $\tilde{\Delta}$ that will also be a proper subset of $[0,1]^n$ for a judicious choice for Δ. The goal is to choose H and Δ such that $\tilde{\Delta} \subset \Delta(F_1, \ldots, F_n)$.

Definition 6.13. Let $\mathcal{H} = \{\phi \in C^0(\mathbb{R}, [0,1]) \,|\, \phi(0) = 0, \phi \text{ is increasing}\}$, where $C^0(\mathbb{R}, [0,1])$ is the set of continuous functions from $\mathbb{R}$ to $[0,1]$, and let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{n-1})$ be in $\mathcal{H}^{n-1} = \underbrace{\mathcal{H} \times \cdots \times \mathcal{H}}_{n-1 \text{ times}}$. The **sub-hypercube domain** $\Delta_{\boldsymbol{\phi}}$ is defined by:

$$\Delta_{\boldsymbol{\phi}} = \{\boldsymbol{x} \in [0,1]^n \,|\, \forall i \in \{1, \ldots, n-1\},\ \phi_i(x_i) \leq x_{i+1}\}$$

and its **lower boundary** by:

$$\partial\Delta_{\boldsymbol{\phi}} = \{\boldsymbol{x} \in \Delta_{\boldsymbol{\phi}} \,|\, \exists i \in \{1, \ldots, n-1\},\ \phi_i(x_i) = x_{i+1}\}$$

Definition 6.14. Let $\Delta_{\boldsymbol{\phi}}$ be a sub-hypercube domain and H be a copula such that $m_{H,\boldsymbol{\phi}} = \mu_H(\Delta_{\boldsymbol{\phi}}) > 0$. The **sub-hypercube distribution function** $D_{H,\boldsymbol{\phi}}$ **associated to** $(H, \boldsymbol{\phi})$ is defined as:

$$\forall B \subset [0,1]^n,\ \mu_{D_{H,\boldsymbol{\phi}}}(B) = \frac{\mu_H(B \cap \Delta_{\boldsymbol{\phi}})}{m_{H,\boldsymbol{\phi}}}$$

Definition 6.15. Let $D_{H,\boldsymbol{\phi}}$ be a sub-hypercube distribution function. The **sub-hypercube copula associated to** $D_{H,\boldsymbol{\phi}}$ is defined as the copula $C_{H,\boldsymbol{\phi}}$ of $D_{H,\boldsymbol{\phi}}$. This copula is uniquely defined as $D_{H,\boldsymbol{\phi}}$ has continuous marginals, and its support $\tilde{\Delta}_{H,\boldsymbol{\phi}}$ is given by:

$$\tilde{\Delta}_{H,\boldsymbol{\phi}} = \left\{\boldsymbol{u} \in [0,1]^n \,|\, \forall i \in \{1, \ldots, n-1\},\ D^{\leftarrow}_{i+1,H,\boldsymbol{\phi}}(u_{i+1}) \geq \phi_i(D^{\leftarrow}_{i,H,\boldsymbol{\phi}}(u_i))\right\}$$

with a lower boundary given by:

$$\partial\tilde{\Delta}_{H,\boldsymbol{\phi}} = \left\{\boldsymbol{u} \in \tilde{\Delta}_{H,\boldsymbol{\phi}} \,|\, \exists i \in \{1, \ldots, n-1\},\ D^{\leftarrow}_{i+1,H,\boldsymbol{\phi}}(u_{i+1}) = \phi_i(D^{\leftarrow}_{i,H,\boldsymbol{\phi}}(u_i))\right\}$$

When H is absolutely continuous with density function h, so are $D_{H,\phi}$ and $C_{H,\phi}$ and the density function $c_{H,\phi}$ of $C_{H,\phi}$ writes:

$$c_{H,\phi}(\boldsymbol{u}) = \frac{h(D^{\leftarrow}_{1,H,\phi}(u_1),\dots,D^{\leftarrow}_{n,H,\phi}(u_n))\mathbb{1}_{\Delta_\phi}(D^{\leftarrow}_{1,H,\phi}(u_1),\dots,D^{\leftarrow}_{n,H,\phi}(u_n))}{m_{H,\phi}\prod_{i=1}^n g_{i,H,\phi}\left(D^{\leftarrow}_{i,H,\phi}(u_i)\right)}$$

where $D_{i,H,\phi}(x_i) = D_{H,\phi}(1,\dots,1,x_i,1,\dots,1)$ is the i-th marginal distribution function of $D_{H,\phi}$ and $g_{i,H,\phi}$ the associated density function.

In a simulation perspective, one may wonder how to generate realizations of such a copula. An obvious way to do it is to use a rejection/transformation technique, as presented in the following algorithm.

Algorithm 6.16. We suppose that we are able to generate realizations of the copula H.

1. Generate $\boldsymbol{v} \in [0,1]^n$ according to H,
2. If $\boldsymbol{v} \notin \Delta_\phi$, go back to 1
3. Compute $\forall i \in \{1,\dots,n\},\ u_i = D_{i,H,\phi}(v_i)$
4. Return $\boldsymbol{u}$

The point $\boldsymbol{v}$ obtained at the end of 2 is distributed to $D_{H,\phi}$ by the rejection technique and the final point $\boldsymbol{u}$ is distributed to $C_{H,\phi}$ thanks to Theorem 1.11. The acceptance ratio is $m_{H,\phi}$. The application of this algorithm in dimension 2 for different choices of copulas H and the same choice of function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\phi(x) = x^2$ gives the realizations illustrated in Figure 6.2.

The rejection rate of this algorithm can be very high, making Algorithm 6.16 very inefficient. We present in section 6.4.2 a rejection-free algorithm (Algorithm 6.26), which is much more efficient in this case.

6.3.2 Properties of the sub-hypercube distribution and copula

The independent copula $\Pi_n(u_1,\dots,u_n) = \prod_{i=1}^n u_i$ is the most entropic copula when one has no information on the dependence structure of a random vector. It motivates us to study the family of sub-square copulas obtained when $H = \Pi_n$.
This section is dedicated to the study of this special case. We simplify the notation with respect to the previous section by dropping the index H as it is implicitly equal to Π_n.
We note by $\mathcal{C}_\phi$ the set of sub-square copulas built with $H = \Pi$.

Proposition 6.17. *In the settings of the previous section, the distribution function of D_ϕ is given by:*

$$\forall \boldsymbol{s} \in [0,1]^n,\ D_\phi(\boldsymbol{s}) = \frac{N(\boldsymbol{s})}{N(\mathbf{1})} \tag{6.10}$$

where $\mathbf{1} = (1,\dots,1)$ and N is given by:

$$N(s_1,\dots,s_n) = \int_0^{s_1} g_{n-1}(x_1, s_2\dots,s_n)\,\mathrm{d}x_1$$

with

$$g_0 = 1$$

and for $k = n-1,\dots,1$:

$$g_{n-k}(x_k, s_{k+1},\dots,s_n) = \mathbb{1}_{\{\phi_k(x_k)\le s_{k+1}\}}\int_{\phi_k(x_k)}^{s_{k+1}} g_{n-k-1}(x_{k+1}, s_{k+2},\dots,s_n)\,\mathrm{d}x_{k+1}$$

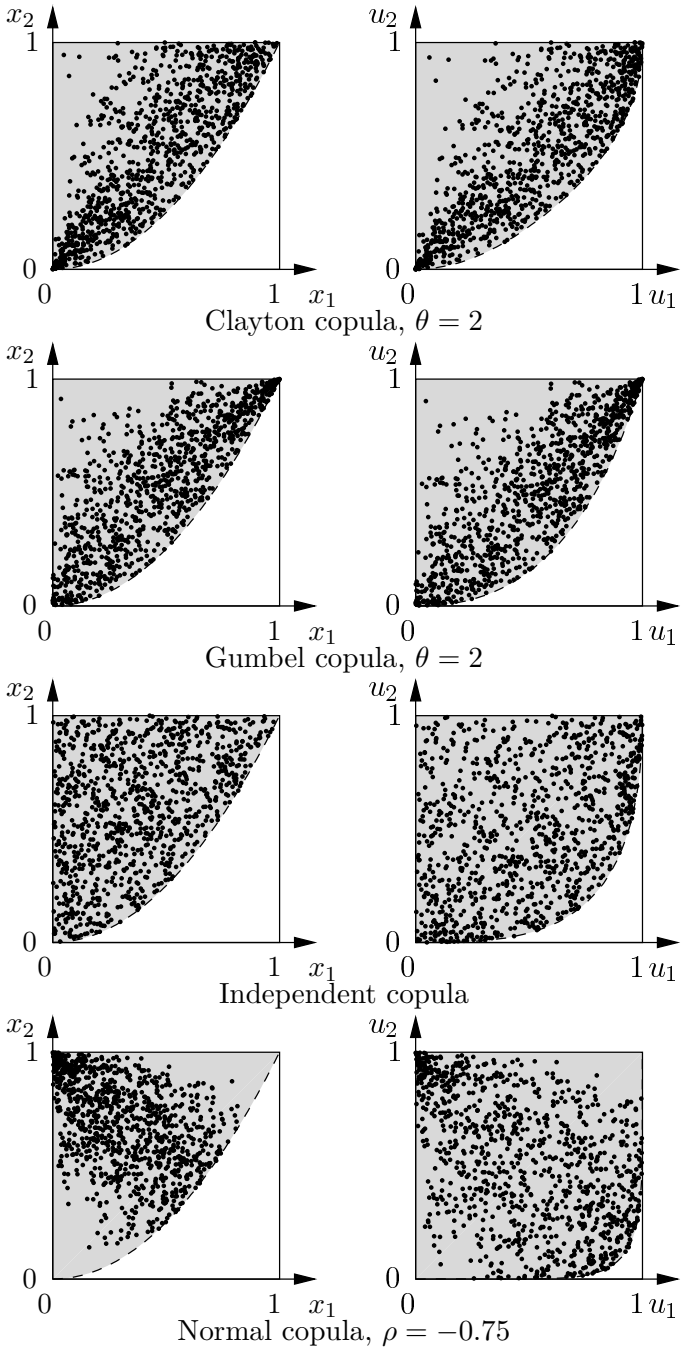


Figure 6.2: 1000 realizations of $D_{H,\phi}$ (left column) and $C_{H,\phi}$ (right column) obtained using Algorithm 6.16, for a given function $\phi(x) = x^2$ and different copulas H. The choice of H modifies the repartition of the points and the support of $C_{H,\phi}$.

Proof. As Π_n is an absolutely continuous copula, the associated sub-hypercube distribution $D_{\boldsymbol{\phi}}$ is also absolutely continuous, with a density function $d_{\boldsymbol{\phi}}$ given by:

$$\forall \boldsymbol{x} \in [0,1]^n,\ d_{\boldsymbol{\phi}}(\boldsymbol{x}) = \frac{\prod_{i=1}^n \mathbb{1}_{\{\phi_i(x_i) \leq x_{i+1}\}}}{\int_0^1 \dots \int_0^1 \prod_{i=1}^n \mathbb{1}_{\{\phi_i(x_i) \leq x_{i+1}\}} \,\mathrm{d}x_1 \dots \mathrm{d}x_n}$$

from which we get the representation (6.10). □

The representation given in Proposition 6.17 allows to compute sequentially both the numerator and the denominator in the expression (6.10) for a given $\boldsymbol{\phi}$, using e.g. a computer algebra system. To give an idea of the increasing complexity of these computations with respect to the dimension n, we give the expression of the normalization factor for $n = 2$ and $n = 3$:

$$\begin{aligned} n = 2, &\quad N(\mathbf{1}) = 1 - \Phi_1(1) \\ n = 3, &\quad N(\mathbf{1}) = 1 - \Phi_1(1) - \Phi_2(1) + \int_0^1 \Phi_2 \circ \phi_1(s)\,\mathrm{d}s \end{aligned}$$

where $\Phi_i(t) = \int_0^t \phi_i(s)\,\mathrm{d}s$.

We will now focus on the bi-dimensional case. We simplify the notation by taking $\phi := \boldsymbol{\phi} = (\phi_1)$ and $\Phi := \Phi_1$.

6.3.3 The bi-dimensional case: sub-square distributions and copulas

If we restrict the analysis to the bi-dimensional case, we are able to explicit all the quantities related to the sub-hypercube distribution and copula, renamed as **sub-square distributions and copulas** in order to avoid confusion with the general case:

Proposition 6.18. *In the settings of the previous section, we have the following properties for D_ϕ:*

- *The mass of Δ_ϕ with respect to Π writes:*

$$m_\phi = 1 - \Phi(1)$$

- *The density function of D_ϕ writes:*

$$g_\phi(\boldsymbol{x}) = \frac{1}{m_\phi} \mathbb{1}_{\{\phi(x_1) \leq x_2\}}$$

- *The distribution function of D_ϕ writes:*

$$\forall \boldsymbol{x} \in [0,1]^2,\ D_\phi(\boldsymbol{x}) = \begin{cases} \dfrac{1}{m_\phi}\left[x_1 x_2 - \Phi(x_1)\right] & \text{on } \Delta_\phi \\ \dfrac{1}{m_\phi}\left[x_2 \phi^{\leftarrow}(x_2) - \Phi \circ \phi^{\leftarrow}(x_2)\right] & \text{on } \Delta_\phi^c \end{cases}$$

- *The marginal distribution functions of D_ϕ write:*

$$\forall x_1 \in [0,1],\ D_{1,\phi}(x_1) = \frac{1}{m_\phi}\left[x_1 - \Phi(x_1)\right] \tag{6.11}$$

$$\forall x_2 \in [0,1],\ D_{2,\phi}(x_2) = \frac{1}{m_\phi}\left[x_2 \phi^{\leftarrow}(x_2) - \Phi \circ \phi^{\leftarrow}(x_2)\right] \tag{6.12}$$

– *The marginal density functions of D_ϕ write:*

$$\forall x_1 \in [0,1],\ d_{1,\phi}(x_1) = \frac{1}{m_\phi}\left[1-\phi(x_1)\right] \tag{6.13}$$

$$\forall x_2 \in [0,1],\ d_{2,\phi}(x_2) = \frac{1}{m_\phi}\left[\phi^{\leftarrow}(x_2)\right] \tag{6.14}$$

– *The distribution function of the sub-square copula C_ϕ writes:*

$$\forall \boldsymbol{u} \in [0,1]^2,\ C_\phi(\boldsymbol{u}) = D_\phi(D^{\leftarrow}_{1,\phi}(u_1), D^{\leftarrow}_{2,\phi}(u_2))$$

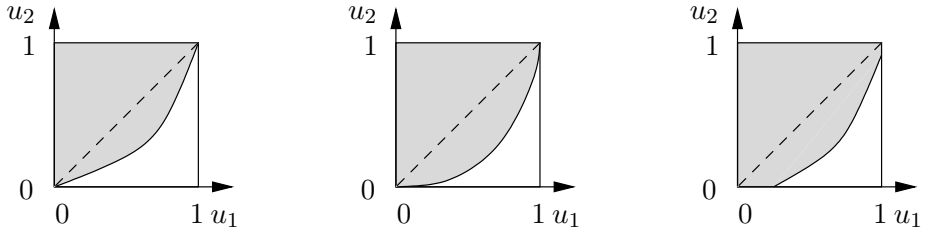
– *The density function of the sub-square copula C_ϕ writes:*

$$\forall \boldsymbol{u} \in [0,1]^2,\ c_{H,\phi}(\boldsymbol{u}) = \frac{m_\phi}{\left[1-\phi\circ D^{\leftarrow}_{1,\phi}(u_1)\right]\left[\phi^{\leftarrow}\circ D^{\leftarrow}_{2,\phi}(u_2)\right]}\mathbb{1}_{\tilde{\Delta}_\phi}(\boldsymbol{u})$$

– *The lower boundary $\partial\tilde{\Delta}_\phi$ can be parametrized by:*

$$\partial\tilde{\Delta}_\phi = \left\{\left(\frac{1}{m_\phi}\left[t-\Phi(t)\right], \frac{1}{m_\phi}\left[t\phi(t)-\Phi(t)\right]\right), \forall t \in [0,1]\right\} \tag{6.15}$$

We introduced the family of sub-square copula to exhibit elements in $\mathcal{C}(F_1,F_2)$ different from the min copula M_2 in the case (c) of Figure 6.1. More precisely, this case can be subdivided into nine sub-cases depending on the slope of the boundary at $(0,0)$, $(1,1)$, the relative positions of the lower boundaries $\underline{X}_1$ and $\underline{X}_2$ and the upper boundaries $\overline{X}_1$ and $\overline{X}_2$. This distinction is illustrated on Figure 6.3, focusing on the behaviour at $(0,0)$.



(c-1) $\underline{X}_1 = \underline{X}_2$, $\rho(0) > 0$ (c-2) $\underline{X}_1 = \underline{X}_2$, $\rho(0) = 0$ (c-3) $\underline{X}_1 < \underline{X}_2$, $\rho(0) = 0$

Figure 6.3: Some of the possible sub-cases of the non-trivial case $\mathcal{C}(F_1,F_2) \subsetneq \mathcal{C}$ corresponding to case (c) on Figure 6.1. The cases (c-2) and (c-3) differ by the fact that in (c-3), the boundary of $\Delta(F_1,F_2)$ has an horizontal part on the left.

Cases (c-1) and (c-2) correspond to $\underline{X}_1 = \underline{X}_2$ with $\rho(0) = \lim\limits_{s\to\underline{X}_1^+}\dfrac{F_2(s)}{F_1(s)} > 0$ in the case (c-1) and $\rho(0) = 0$ in the case (c-2). The case (c-3) corresponds to $\underline{X}_1 < \underline{X}_2$, which means that the boundary $\partial\Delta(F_1,F_2)$ starts by an horizontal segment. The distinction between cases (c-2) and (c-3) will play a central role in the study of the sub-square copulas.

Using the relation (6.15), we can study the slope of the lower boundary in the vicinity of $t=0$ or $t=1$. In the vicinity of $t=0$, the slope $\rho(0)$ writes:

$$\rho(0) = \lim_{t\to 0^+}\frac{t\phi(t)-\Phi(t)}{t-\Phi(t)}$$

As $\phi(0) = 0$ and $\Phi(0) = 0$, $\Phi(t) = \Phi(0) + t\phi(t) + o(t) = o(t)$, so $\rho(0) = 0$, which shows that no sub-square copula is admissible in the case (c-1) of Figure 6.3. We will now focus on the case where:

$$\underline{X}_1 < \underline{X}_2 < \overline{X}_1 < \overline{X}_2 \tag{6.16}$$

and introduce $\alpha = F_1(\underline{X}_2) \in (0,1)$ and $\beta = F_2(\overline{X}_1) \in (0,1)$.

The main result of the next section is to prove that in this case, there exists a unique sub-square copula with maximal support $\Delta(F_1, F_2)$.

6.4 Characterization of compatible sub-square copulas with largest possible support

In this part, under the hypothesis (6.16), we characterize the functions ϕ such that the resulting copula C_ϕ has its support $\tilde{\Delta}_\phi$ such that $\tilde{\Delta}_\phi = \Delta(F_1, F_2)$, which is the largest support for copula in $\mathcal{C}(F_1, F_2)$.

6.4.1 Existence and uniqueness of a maximal sub-square copula

This characterization is based on the equality between the boundaries of Δ_ϕ and $\Delta(F_1, F_2)$. In the case where $\underline{X}_2 < \overline{X}_1$, which is the situation we focus on, this boundary is given by:

$$\partial\Delta(F_1, F_2) = \left\{ (F_1(s), F_2(s)), \quad s \in [\underline{X}_1, \overline{X_1}] \right\} \tag{6.17}$$

which encompasses two parts: an horizontal segment for $s \in [\underline{X}_1, \underline{X}_2]$ as $F_2(s) = 0$, and a curve for the other values of s. Using relations (6.15) and (6.17), the increasing properties of $G_{i,\phi}$, F_i and the relations $F_i \circ F_i^{\leftarrow} = Id$ for $i = 1, 2$, we can rewrite the equality of the two lower bounds of the supports as:

$$\begin{cases} t\phi(t) & = & \Phi(t) + m_\Phi J\left(\frac{t-\Phi(t)}{m_\Phi}\right) & \forall t \in [0,1] \\ \Phi(1) & = & 1 - m_\Phi & \\ \Phi(\alpha m_\Phi) & = & 0 & \end{cases} \tag{6.18}$$

where $J = F_2 \circ F_1^{\leftarrow}$.

In order to solve (6.18), we introduce the following parametric problem with respect to $m \in (0,1)$ defined for $t \in (0, +\infty)$:

$$\begin{cases} \Phi'_m(t) & = & \frac{1}{t}\Phi_m(t) + \frac{m}{t} J\left(\frac{t-\Phi_m(t)}{m}\right) & \forall t \in (0, +\infty) \\ \Phi_m(1) & = & 1 - m & \end{cases} \tag{6.19}$$

and we look for m such that

$$\Phi_m(\alpha m) = 0 \text{ and } \Phi_m \text{ continuous at 0 with } \Phi_m(0) = 0 \tag{6.20}$$

The solutions of (6.18) are the restrictions to $[0,1]$ of the solutions of (6.19) that verify (6.20) and that can be extended to 0 by continuity. For $a \in (0,1)$ and $b \in (a,1]$ we define $\mathcal{J}_{a,b}$ as:

$$\begin{aligned} \mathcal{J}_{a,b} = \{ J \in C^0(\mathbb{R}, \mathbb{R}) \,|\, & J \text{ is increasing}, \forall x \le a, J(x) = 0, \\ & \forall x \ge 1, J(x) = b, \forall x \ge 0, J(x) \le x \} \end{aligned} \tag{6.21}$$

where $C^0(\mathbb{R},\mathbb{R})$ is the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

One can easily check that for any continuous univariate distribution functions F_1 and F_2 with F_1 invertible, then $F_2 \circ F_1^{-1} \in \mathcal{J}_{\alpha,\beta}$ where α and β are given in (6.16). The following proposition shows that the converse is also true:

Proposition 6.19. *Let J be in $\mathcal{J}_{a,b}$. Then there exist continuous univariate distribution functions F_1, F_2 with F_1 invertible such that $J = F_2 \circ F_1^{-1}$.*

Proof. It is enough to take for F_1 the distribution function of $\mathcal{U}(0,1)$ and to define F_2 as $F_2(x) = J(x)$ for all $x \leq 1$ and to complete it on $[1,+\infty)$ such that the resulting F_2 is continuous, increasing and $\lim_{s\to+\infty} F_2(x) = 1$. □

Let v be defined by

$$\forall s \in \mathbb{R},\ v(s) = \frac{e^{-s} - \Phi_m(e^{-s})}{m} \tag{6.22}$$

Then, using (6.19), v satisfies:

$$\begin{cases} v'(s) &= -v(s) + J\left(v(s)\right) \quad \forall s \in \mathbb{R} \\ v(0) &= 1 \end{cases} \tag{6.23}$$

which is an autonomous ordinary differential equation which does not depend on m.

Remark 6.20. For a given $m \in (0,1)$, the transformation $(t, \phi_m(t)) \to (s = -\log t, v(s))$ defined by equation (6.22) is a diffeomorphism. It ensures that the solutions of equation (6.19) are in one-to-one correspondence with the solutions of equation (6.23).

Remark 6.21. As $\forall x \in \mathbb{R},\ J(x) \leq x$, any solution v of (6.23) is non-increasing on its definition domain.

Remark 6.22. As $\forall x \in \mathbb{R},\ 0 \leq J(x) \leq b$, then any solution v of (6.23) verifies $\forall s \geq 0, e^{-s} \leq v(s) \leq b + (1-b)e^{-s}$ and $\forall s \leq 0, b + (1-b)e^{-s} \leq v(s) \leq e^{-s}$.

Theorem 6.23. *Let J be in $\mathcal{J}_{a,b}$ with $a \in (0,1)$ and $b \in (a,1]$. If J is k-lipschitz on $[a,1]$, then (6.23) has a unique global solution v defined on $\mathbb{R}$. This solution is such that $\lim_{s\to+\infty} v(s) = x^*$, where $x^* = \max\{x \in [0,1] \,|\, J(x) = x\}$.*
Furthermore, the equation $v(s) = a$ has a solution s^* if and only if $x^* = 0$. In this case, s^* is unique and verifies $s^* > -\log a$, and $\forall s \geq s^*,\ v(s) = \alpha e^{-(s-s^*)}$.

Proof. We first note that $\{x \in [0,1] \,|\, J(x) = x\}$ is closed, and non-empty as $J(0) = 0$. It insures that x^* is well defined.
The hypotheses made on J give that J is k-lipschitz on $\mathbb{R}$. Using the Cauchy-Lipschitz theorem, there exists a unique maximal solution of (6.23) defined on $I \subset \mathbb{R}$. If v were not a global solution, it would escape any compact subset of $\mathbb{R}$ in finite time, which is incompatible with the bounds (6.22), so $I = \mathbb{R}$.

Remarks 6.21 and 6.22 show that $\ell = \lim_{s\to+\infty} v(s)$ exists and for all $s \in \mathbb{R}$, $v(s) \geq \ell \geq 0$. Suppose that $\ell > x^*$, then by definition of x^*, $\eta = \inf_{[\ell,1]}(x - J(x)) > 0$, and for all $s \geq 0$, $v'(s) \leq -\eta$.It implies that $v(s) \leq 1 - \eta s < x^*$ for $s > \frac{1-x^*}{\eta}$, which contradicts $v(s) \geq \ell > x^*$, so $\ell = x^*$. If $x^* > 0$, as for all $x \in (0,a]$, $J(x) = 0 < x$, we have $x^* > a$. In this case, for all $s \geq 0$, $v(s) \geq x^* > a$ and there is no s such that $v(s) = a$. If $x^* = 0$, as $v(0) = 1 > a$ and $\lim_{s\to+\infty} v(s) = 0 < a$, by continuity $S_a = \{s \geq 0 \,|\, v(s) < a\} \neq \emptyset$ and $s^* = \inf S_a > 0$ is such that $v(s^*) = a$ and for all $s \in [0, s^*)$, $v(s) > a$. We deduce that $v'(s^*) \leq -\inf_{[a,1]}(x - J(x)) < 0$, and using Remark 6.21, for all $s > s^*$, $v(s) < a$. We conclude that s^* is the unique solution of $v(s) = a$.

To show that $s^* > -\log a$, we refine the lower bound in Remark 6.22) for $s > 0$. We have $J(1) = b > 0$. By continuity, there exists $\epsilon > 0$ such that for all $x \in [1-\epsilon, 1]$, $J(x) \geq b/2 > 0$. As we also have $v(0) = 1$, by continuity, there exists $\eta > 0$ such that for all $s \in [0, \eta]$, $v(s) \in [1-\epsilon, 1]$. Then, for all $s \in [0, \eta]$, $v'(s) \geq -v(s) + b/2$.

Let define u_η as the solution of $u_\eta(0) = 1$ and for all $s \in [0, \eta]$, $u'_\eta(s) = -u_\eta(s) + b/2$. Then for all $s \in (0, \eta]$, $u_\eta(s) = (1 - b/2)e^{-s} + b/2$ and $v(s) \geq u_\eta(s) > e^{-s}$.
Let define v_η as the solution of $v_\eta(\eta) = u_\eta(\eta)$ and for all $s \in [\eta, +\infty)$, $v'_\eta(s) = -v_\eta(s)$. Then $v_\eta(s) = [1 + b/2(e^\eta - 1)]e^{-s}$ and for all $s \in [\eta, +\infty)$, $v(s) \geq v_\eta(s) > e^{-s}$.

We conclude that for all $s > 0$, $v(s) > e^{-s}$. In particular, $v(-\log a) > a$ which shows that $s^* > -\log a$. For $s \geq s^*$, v is such that $v(s^*) = \alpha$ and $v'(s) = -v(s)$ so $v(s) = \alpha e^{-(s-s^*)}$. □

Corollary 6.24. *Let F_1 and F_2 be continuous univariate distribution functions verifying the compatibility condition (6.3) and the hypothesis (6.16). If $J = F_2 \circ F_1^{\leftarrow} \in \mathcal{J}_{\alpha,\beta}$ is k-lipschitz on $[\alpha, 1]$, then there exists a sub-square copula $C_\phi \in \mathcal{C}(F_1, F_2)$ if and only if for all $x \in (\underline{X}_2, \overline{X}_1]$, $F_1(x) > F_2(x)$. In this case, C_ϕ is unique and ϕ is given by:*

$$\begin{cases} \phi(0) &= 0 \\ \phi(t) &= 1 + \frac{e^{-s^*}}{\alpha t}\left(J(v(-\log t)) - v(-\log t)\right) \quad \forall t \in (0, 1] \end{cases}$$

where v is the solution (6.23). We note that for all $t \in [0, e^{-s^*}]$, $\phi(t) = 0$.

Proof. The existence of a sub-square copula C_ϕ is equivalent to the existence of a pair (m, Φ_m) where Φ_m is a solution of equation (6.19) that satisfies the constraint (6.20). By Remark 6.20 the existence of ϕ_m for a given m is equivalent to the existence of a solution v. Using (6.22), the constraint (6.20) translates into the existence of $s^* > -\log\alpha$ such that $v(s^*) = \alpha$ and the existence of $\ell = \lim_{s\to+\infty} v(s)$ such that $\ell = 0$. It results that the existence and the uniqueness of a solution v of (6.23) with such a s^* and ℓ is equivalent to the existence and uniqueness of C_ϕ.

The hypotheses made on F_1, F_2 and J ensure that the solution v of (6.23) exists and is unique, and the condition $\forall x \in (\underline{X}_2, \overline{X}_1)$, $F_1(x) > F_2(x)$ translates into $x^* = \max\{x \in [0, 1] \,|\, J(x) = x\} = 0$. By Proposition 6.23, it is equivalent to the existence and uniqueness of a sub-square copula $C_\phi \in \mathcal{C}(F_1, F_2)$.

Using (6.22) and (6.23), we have $m^* = \frac{e^{-s^*}}{\alpha} \in (0, 1)$ and $\phi_{m^*} = \Phi'_{m^*}$ is given by:

$$\phi_{m^*}(t) = 1 + \frac{e^{-s^*}}{\alpha t}\left(J(v(-\log t)) - v(-\log t)\right) \quad \forall t \in (0, 1]$$

For $s \geq s^*$, the expression of $v(s)$ shows that $\forall t \in (0, \alpha m^*]$, $\Phi_{m^*}(t) = 0$. As $\ell = 0$, Φ_{m^*} can be continuously extended to 0 by setting $\Phi_m(0) = -\ell m^* = 0$. It results that Φ_{m^*} is null on $[0, \alpha m^*] = [0, e^{-s^*}]$, so is ϕ_{m^*}. □

6.4.2 Numerical aspects

It remains to translate the preceding results into a simulation procedure. A straightforward approach is presented in Algorithm 6.25.

Algorithm 6.25. Under the hypotheses of Corollary 6.24, the following procedure builds a discretization of ϕ, the parameter of the unique maximal sub-square copula $C_\phi \in \mathcal{C}(F_1, F_2)$:

1. Compute a lower bound $\tilde{\eta} \leq \eta$ of $\eta = \min_{[\alpha,1]}(t - J(t))$
2. Define a regular grid $(s_k)_{k\in\{0,\dots,N-1\}}$ on $[0, 1/\tilde{\eta}]$ such that $s_0 = 0$ and $s_{N-1} = 1/\tilde{\eta}$ and the associated grid $(t_k)_{k\in\{0,\dots,N-1\}}$ such that $t_k = \exp(-s_{N-1-k})$.
3. Solve numerically (6.23) using $(s_k)_{k\in\{0,\dots,N-1\}}$. It gives the pairs $(s_k, \widehat{v}_k)$ as an approximation of $(s_k, v(s_k))$.
4. Compute an approximate value $\widehat{s}^*$ of s^* e.g. by linear interpolation in $(s_k, \widehat{v}_k)$: if k^* is the largest index such that $\widehat{v}_{k^*} > \alpha$, $\widehat{s}^* = s_{k^*} + (\alpha - \widehat{v}_{k^*}) \left(\frac{s_{k^*+1} - s_{k^*}}{\widehat{v}_{k^*+1} - \widehat{v}_{k^*}} \right)$. As this step is crucial for the quality of the resulting approximation, the linear interpolation may be too crude. A possible alternative is to use a bisection approach: starting from the bracketing interval $[s_{k^*}, s_{k^*+1}]$, we compute $\widehat{v}_{k^*+1/2} \simeq v(s_{k^*+1/2})$ using one step of the integration method with a step of length $\frac{s_{k^*+1} - s_{k^*}}{2}$. We update the bounds of the bracketing interval depending on the position of $\widehat{v}_{k^*+1/2}$ with respect to α.
5. Using $\widehat{s}^*$ and $(s_k, \widehat{v}_k)$, build an approximation $(t_k, \widehat{\phi}_{\widehat{m}^*,k})$ of $(t_k, \phi_{m^*}(t_k))$ with $\widehat{m}^* = \frac{e^{-\widehat{s}^*}}{\alpha}$ and $\widehat{\phi}_{\widehat{m}^*,k} = 1 + \frac{\widehat{m}^*}{t_k} \left(J(\widehat{v}_{N-k+1}) - \widehat{v}_{N-k+1} \right)$
6. A continuous approximation $\widehat{\phi}$ of ϕ is obtained as:

$$\widehat{\phi}(t) = \begin{cases} 0 & \forall t \in [0, t^*] \\ \left(\frac{t - t^*}{t_{k^*} - t^*} \right) \widehat{\phi}_{k^*} & \forall t \in [t^*, t_{k^*}] \\ \widehat{\phi}_k + \left(\frac{t - t_k}{t_{k+1} - t_k} \right) (\widehat{\phi}_{k+1} - \widehat{\phi}_k) & \forall t \in [t_k, t_{k+1}] \text{ and } k \in \{k^* + 1, \dots, N-1\} \end{cases}$$

For the numerical experiments, we took a non-adaptive fixed-step fourth order Runge-Kutta method for the step 3 of the algorithm and a bisection method for the step 4.

Once the piecewise linear continuous approximation $\widehat{\phi}$ is obtained, it is possible to compute all the functions associated to $C_{\widehat{\phi}}$, such as its density or its distribution function. As $\Delta_{\widehat{\phi}}$ is a polygon, as it can be seen on Figure 6.4, it is possible to sample with respect to the sub-square distribution $D_{\widehat{\phi}}$ without rejection, thanks to Algorithm 6.26. The main cost of Algorithm 6.16 is thus avoided, and the resulting random generator for $C_{\widehat{\phi}}$ is efficient.

Algorithm 6.26. The distribution $D_{\widehat{\phi}}$ can be written as a discrete mixture of uniform distributions over the triangles T_0, $T_{k^*}, \dots, T_{N+1}$, the weights w_i of the mixture being proportional to the surface of these triangles:

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \sum_{i \in I} \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2 \,|\, (X_1, X_2) \in T_i) \mathbb{P}((X_1, X_2) \in T_i)$$

where $I = \{0, k^*, \dots, N+1\}$.

1. Generate k, a realization of an integer-valued random variable K distributed according to the discrete distribution (w_i, I) with $w_i = \mathbb{P}((X_1, X_2) \in T_i)$, i.e. $\forall i \in I$, $\mathbb{P}(K = i) = w_i$. This step can be done in $\mathcal{O}(1)$ time using e.g. the method of aliases.
2. Generate u and v, realizations of respectively U and V, two random variables uniformly distributed over $[0, 1]$. The variables U, V and K are independent.

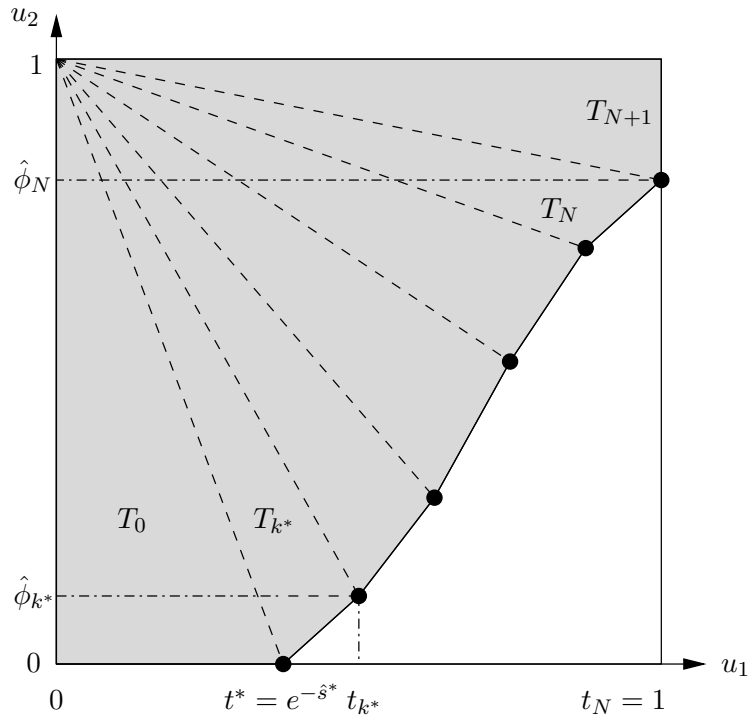


Figure 6.4: Graph of the piecewise function $\widehat{\phi}$ and the associated support of $D_{\widehat{\phi}}$. The polygonal nature of this support allows to sample $D_{\widehat{\phi}}$ very efficiently, using Algorithm 6.26

3. Return the weighted average of the vertexes $\boldsymbol{a}_1^k$, $\boldsymbol{a}_2^k$, $\boldsymbol{a}_3^k$ of the triangle T_k with weights $(\min(u,v), |v-u|, 1-\max(u,v))$. The vertices are given by:

$$\boldsymbol{a}_1^i = \quad (0,1) \quad i = 0, k^*, \ldots, N+1$$

$$\boldsymbol{a}_2^i = \begin{cases} (t^*, 0) & i = 0 \\ (t_i, \widehat{\phi}_i) & i = k^*, \ldots, N \\ (1,1) & i = N+1 \end{cases}$$

$$\boldsymbol{a}_3^i = \begin{cases} (0,0) & i = 0 \\ (t^*, 0) & i = k^* \\ \boldsymbol{a}_2^{i-1} & i = k^*+1, \ldots, N+1 \end{cases}$$

Example 6.27. Let F_1 be the distribution function of $\mathcal{U}[0,1]$ and F_2 of $\mathcal{U}[a, a+1]$ with $0 \leq a \leq 1$. The several functions involved in the construction of the associated sub-square copula are:

$$J(t) = \begin{cases} 0 & \forall t \in [0, a) \\ t - a & \forall t \in [a, 1] \end{cases}$$

$$v(s) = \begin{cases} 1 - as & \forall s \in [0, s^*) \\ a \exp(-s + 1/a - 1) & \forall s \in [s^*, +\infty) \end{cases}$$

with $s^* = 1/a - 1$

$$\phi(t) = \begin{cases} 0 & \forall t \in [0, e^{-s^*}) \\ 1 - \dfrac{e^{-s^*}}{t} & \forall t \in [e^{-s^*}, 1] \end{cases}$$

The application of Algorithm 6.25 gives the results presented on Figure 6.5, for the cases $a \in \{2^{-n} \,|\, n = 1, \dots, 5\}$. The following numerical settings are used:

1. The value of $\widehat{\eta}$ is obtained as the minimum value of $t_k - J(t_k)$, where k is such that $t_k \in \left\{\frac{\lceil aN \rceil}{N-1}, \dots, 1\right\}$. Here, $N = 10^4$. For the given example, we have $\widehat{\eta} = \eta$.
2. The grid is built using the same value N as in the previous step.
3. The numerical integration is done using a fixed-step fourth order Runge-Kutta method with N points. It gives an approximation of v which is exact up to machine precision for $s \leq s^*$ as v is linear on this interval.
4. The value of s^* is computed using the bisection method, up to machine precision.

The sub-square copula is built using the piecewise linear approximation of ϕ as described in Algorithm 6.25, in order to use Algorithm 6.26 for its simulation. The value of $\widehat{m}^*$ obtained from $\widehat{s}^*$ is not the value associated to the piecewise linear approximation of ϕ obtained at the last step of the algorithm. The quality of this approximation can be assessed thanks to the relative error made on m^* using the exact value $\widetilde{m}^*$ associated to the piecewise approximation. It leads to the quality measure $\epsilon_{m^*} = \frac{|m^* - \widetilde{m}^*|}{m^*} \simeq \frac{|\widehat{m}^* - \widetilde{m}^*|}{\widehat{m}^*}$.

As seen in Table 6.2, for small values of a, the associated mass m^* is so small that Algorithm 6.16 becomes unusable, justifying the use of Algorithm 6.26 that does not suffer from this mass reduction. For example, with $a = 1/32$, $m^* = 32/e^{31} \simeq 1.10\,10^{-12}$, which means that only about one realization over one thousand billions is accepted in step 2 of Algorithm 6.16!

a	$1/2$	$1/4$	$1/8$	$1/16$	$1/32$
s^*	1	3	7	15	31
m^*	0.736	0.199	7.30×10^{-3}	4.89×10^{-6}	1.10×10^{-12}
ϵ_{m^*}	3×10^{-9}	2×10^{-8}	9×10^{-8}	4×10^{-7}	2×10^{-6}

Table 6.2: Numerical results of Algorithm 6.25 for the selected values of a. The relative precision of the algorithm is quantified by $\epsilon_{m^*} = \frac{|m^* - \widetilde{m}^*|}{m^*}$. The values of $\widetilde{m}^*$ are not given, as it is equal to the value of m^* to at least 5 digits for all the values of a.

6.5 Conclusion

In this chapter, we fully characterized multivariate distributions of order statistics with continuous marginal distribution functions, first in terms of marginal distribution functions, then in terms of copulas.

We showed that excepted in the trivial case of deterministic ordering of the supports of the marginal distributions, for which any copula can be used to get a joint distribution function that satisfies the ordering constraints, none of the classical copulas are compatible with the constraints. Then, we proposed a generic construction of compatible copulas, called sub-hypercube copulas, and fully characterized these copulas in the bivariate case. We also provided all the algorithmic details to build and sample such a copula given compatible marginal distribution functions.

Several extensions to the full multi-dimensional case are possible. The first one is to explore in more details the general sub-hypercube copulas. The second one is to use the

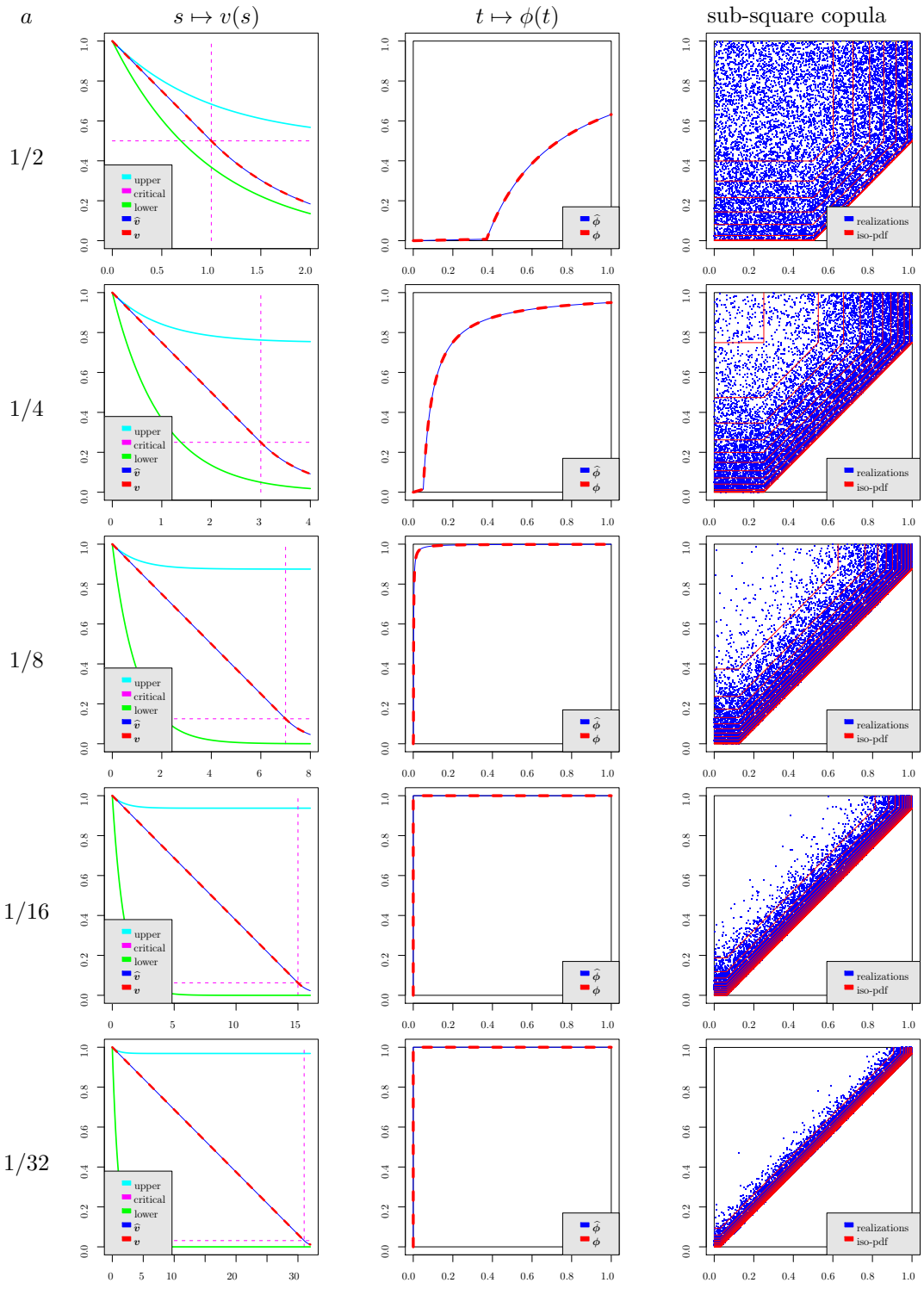


Figure 6.5: Computation of the maximal sub-square copula for the uniform/uniform case, for various values of a. We see graphically the convergence of the sub-square copula to the min copula M_2 when $a \to 0^+$, as $\Delta(F_1, F_2) \to \mathcal{S}_U$.

sub-square copulas in a vine pair-copulas approach, which is a tree-based association of bi-dimensional copulas to build compatible multi-dimensional copulas. See [KC06, Chapter 4] for an introdution to high dimensional dependence modeling using vine copulas.

Chapter 7

Multi-Dimensional discrete distributions

In the continuous case, the copula theory is perfectly suited to the modeling of stochastic dependence as there is a one-to-one correspondence between the joint distribution function and the set of marginal distributions and the copula. In the discrete case, it is no longer the case, and many different copulas can be used to get exactly the same joint distribution function. As shown in [GN07], this lack of uniqueness can be the root of many mistakes when transposing practices from the continuous case to the discrete one, and a direct approach to the stochastic modeling of discrete models using the joint distribution function is sometimes the best way to go.

The key result of this chapter is the presentation of an original and very efficient algorithm to compute such a joint distribution function. It allows to deal with problem dimensions that were out of reach using the previously published algorithms for an actual accuracy that matches or even outperforms them.

This work has been published in [Leb12].

7.1 Introduction

The computation of rectangular probabilities of multi-dimensional discrete integer distributions such as the multinomial, multi-dimensional hypergeometric or multi-dimensional Pólya distributions is of great interest both for statistical applications and for probabilistic modeling purpose. All these distributions are members of a family of multi-dimensional discrete integer distributions for which the existing methods to evaluate such probabilities are either approximate, with no real control on the precision of the approximation, or exact (if the computation is made using exact arithmetic) but available only for some of these distributions or for particular rectangular probabilities.

We propose here a new approximate algorithm that allows performing these computations in the most general case for both the distribution and the rectangular region. Its accuracy matches or even outperforms the exact algorithms when the rounding errors are taken into account. In the worst case, the computational cost of our algorithm is the same as the most efficient exact method published so far, and is much lower in many situations of interest. Our algorithm does not need an intermediate storage that grows with the dimension or problem parameters, which allows dealing with large dimension/large counting parameter applications at no memory cost and acceptable computation time, which is a major difference with respect to the methods published so far.

We are interested in the computation of rectangular probabilities for a d-dimensional discrete integer-valued random vector $\boldsymbol{X} \sim \mathcal{D}$:

$$\begin{aligned} &\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{N}^d, \\ &\quad p_{\mathcal{D}}(\boldsymbol{a}, \boldsymbol{b}) &&= \mathbb{P}\left(\boldsymbol{a} \leq \boldsymbol{X} \leq \boldsymbol{b}\right) \\ &&&= \mathbb{P}\left(a_1 \leq X_1 \leq b_1, \ldots, a_d \leq X_d \leq b_d\right) \end{aligned} \tag{7.1}$$

The computation of such quantities are of uttermost interest in many statistical applications for $\boldsymbol{X}$ distributed according to a multinomial, multi-dimensional hypergeometric or multi-dimensional Pólya distribution (see [BS98], [Cor11], [Fre09], [Goo57], [Joh60], [Lev81], [Lev83], [Lev92]), but despite the existing literature on the subject, no function allows to perform this computation in the standard numerical softwares such as R, SAS, Matlab, Scilab or Octave.

Several authors (see [BS98], [CB00], [Goo57], [Lev81], [Lev83], [Lev92], [Mal68]) have described in details approximate algorithms for a long time, but these algorithms provide only a limited precision which may be inadequate for some applications, and with no control on the error. This lack of control of the error may be the reason why these algorithms have not yet been implemented in a standard numerical package. Some of these authors (see [Goo57], [Lev81], [Lev83], [Lev92]) have also indicated how to derive an exact algorithm if one were able to compute a particular convolution exactly, but with no indication on how to do it efficiently and accurately.

It is only recently that reasonably efficient algorithms for the computation of rectangular probabilities have been described (see [Cor11], [Fre09]), using completely different roots than the previous authors. But even with these algorithms, the only case covered with full generality (see [Fre09]) is the multinomial one. The efficiency of an algorithm is measured by two quantities:

- Its **space complexity**, which is a measure of how much storage the algorithm needs to run;
- Its **time complexity**, which is a measure of how much time the algorithm needs to complete. This complexity is often measured in terms of the number of elementary operations the algorithm needs to complete, assuming that all the elementary operations (multiplication, addition, exponentiation, ...) take the same execution time.

These complexities can be measured either in the worst case or in the mean case, assuming a given distribution of the inputs. In our comparison, we will use worst cases complexities. In the case of the algorithms described in [Cor11] and [Fre09], both the space and time complexities are polynomials, it means bounded by polynomial expressions in the parameters of the distribution of interest and the size of the problem, i.e the magnitude of $\boldsymbol{a}$ and $\boldsymbol{b}$.

We list the available algorithms, the distributions they adress and the possible restrictions on $\boldsymbol{a}$ and $\boldsymbol{b}$ in the computation of (7.1) in Table 7.1 for the exact algorithms and in Table 7.2 for the approximate ones.

We propose to change this situation by providing an algorithm which is essentially exact up to machine precision for all the multi-dimensional discrete distributions considered in [BS98] and [Lev83], which include the multinomial, multi-dimensional hypergeometric and multi-dimensional Pólya distributions. In the multinomial case, our algorithm is more efficient with respect to both space and time complexity than the algorithm described in [Fre09], for an equivalent accuracy when implemented in double precision.

More precisely, we are interested in d-dimensional discrete distributions $\mathcal{D}$ with a $(d-1)$-dimensional probability function. We suppose that there exists a random vector $\boldsymbol{Y}_t =$

$(Y_{t1}, \dots, Y_{td})$ with independent components such that $\boldsymbol{X} \sim \mathcal{D}$ has the same distribution as $\boldsymbol{Y}_t \,|\, \sum_{j=1}^d Y_{tj} = N$, where $t > 0$ is a scaling parameter for the mean of $\boldsymbol{Y}_t$.

With these hypotheses, the rectangular probability (7.1) admits the following representation by a direct application of Bayes' theorem:

$$p_{\mathcal{D}}(\boldsymbol{a}, \boldsymbol{b}) = \mathbb{P}(T_t = N) \frac{\prod_{j=1}^d \mathbb{P}(a_j \leq Y_{tj} \leq b_j)}{\mathbb{P}(Y_t = N)} \tag{7.2}$$

where

$$T_{tj} = (Y_{tj} | a_j \leq Y_{tj} \leq b_j), \; T_t = \sum_{j=1}^d T_{tj} \text{ and } Y_t = \sum_{j=1}^d Y_{tj} \tag{7.3}$$

We also suppose that all the variables Y_{tj} are members of a parametric family of distributions $\mathcal{L}(\boldsymbol{\theta})$ for which the distribution of Y_t is known analytically. It is the case if $\mathcal{L}(\boldsymbol{\theta})$ is closed under convolution, i.e. $\forall j \in \{1, \dots, d\}$, $Y_{tj} \sim \mathcal{L}(\boldsymbol{\theta}_j)$ and $Y_t \sim \mathcal{L}(\boldsymbol{\theta})$. This last hypothesis is not essential: if it is not fulfilled, then one can evaluate $\mathbb{P}(Y_t = N)$ using the same method as the one used to evaluate $\mathbb{P}(T_t = N)$.

This set of hypotheses cover the multinomial, multi-dimensional hypergeometric and multi-dimensional Pólya distributions. If we remove the last hypothesis, we can also include the non-central multi-dimensional hypergeometric and the non-central negative multi-dimensional hypergeometric distributions also called the non-central multi-dimensional Pòlya distribution, as defined in [Ma99].

We recall some definitions concerning these distributions and make explicit the associated family $\mathcal{L}(\boldsymbol{\theta})$. We note $\mathcal{S}$ the set $\{\boldsymbol{a} \in \mathbb{N}^d | \sum_{j=1}^d a_j = N\}$, and we have:

Definition 7.1. The multinomial distribution $\mathcal{M}_d(N, \boldsymbol{p})$ is defined by:

$$\begin{aligned} &\forall \boldsymbol{x} \in \mathbb{N}^d, \\ \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \quad &= \quad \frac{N!}{\prod_{j=1}^d x_j!} \left(\prod_{j=1}^d p_j^{x_j} \right) \mathbb{1}_{\mathcal{S}}(\boldsymbol{x}) \end{aligned} \tag{7.4}$$

where $\forall j \in \{1, \dots, d\}, p_j \geq 0$ and $\sum_{j=1}^d p_j = 1$.

The decomposition (7.2) is obtained with

$$\mathcal{L}(\boldsymbol{\theta}_j) = \mathcal{P}(tp_j)$$

where $\mathcal{P}(tp_j)$ is the Poisson distribution with mean tp_j for any $t > 0$, and $Y_t \sim \mathcal{P}(t)$.

Definition 7.2. The multi-dimensional hypergeometric distribution $\mathcal{H}_d(N, \boldsymbol{h})$ is defined by:

$$\begin{aligned} &\forall \boldsymbol{x} \in \mathbb{N}^d, \\ \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \quad &= \quad \left(\prod_{j=1}^d \binom{h_j}{x_j} \Big/ \binom{h}{N} \right) \mathbb{1}_{\mathcal{S}'}(\boldsymbol{x}) \end{aligned} \tag{7.5}$$

where $\forall j \in \{1, \dots, d\}, h_j \in \mathbb{N}$, $\mathcal{S}' = \mathcal{S} \cap \{0, \dots, h_1\} \times \cdots \times \{0, \dots, h_d\}$ and $h = \sum_{j=1}^d h_j$.

The decomposition (7.2) is obtained with

$$\mathcal{L}(\boldsymbol{\theta}_j) = \mathcal{B}(h_j, t)$$

the binomial distribution with mean th_j for any $t \in (0, 1)$, and $Y_t \sim \mathcal{B}(h, t)$.

Definition 7.3. The multi-dimensional Pólya distribution $\mathcal{P}_d(N, \boldsymbol{q})$ is defined by:

$$\begin{aligned} &\forall \boldsymbol{x} \in \mathbb{N}^d, \\ \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) &= \frac{\Gamma(q)N!}{\Gamma(N+q)} \left(\prod_{j=1}^{d} \frac{\Gamma(x_j + q_j)}{\Gamma(q_j) x_j!} \right) \mathbb{1}_{\mathcal{S}}(\boldsymbol{x}) \end{aligned} \tag{7.6}$$

where $\forall j \in \{1, \ldots, d\}, q_j > 0$ and $q = \sum_{j=1}^{d} q_j$.

The decomposition (7.2) is obtained with

$$\mathcal{L}(\boldsymbol{\theta}_j) = \mathcal{NB}(q_j, t)$$

the negative binomial distribution with mean $\frac{q_j(1-t)}{t}$ for any $t \in (0, 1)$, and $Y_t \sim \mathcal{NB}(q, t)$.

and for the two distributions for which Y_t is not in the parametric family of Y_{tj}:

Definition 7.4. The multi-dimensional non-central hypergeometric distribution $\mathcal{H}_d^*(N, \boldsymbol{h}, \boldsymbol{\theta})$ is defined by:

$$\begin{aligned} &\forall \boldsymbol{x} \in \mathbb{N}^d, \\ \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) &= \left(\prod_{j=1}^{d} \binom{h_j}{x_j} \theta_j^{x_j} / C(\boldsymbol{\theta}) \right) \mathbb{1}_{\mathcal{S}'}(\boldsymbol{x}) \end{aligned} \tag{7.7}$$

where $\forall j \in \{1, \ldots, d\}, h_j \in \mathbb{N}$, $\mathcal{S}' = \mathcal{S} \cap \{0, \ldots, h_1\} \times \cdots \times \{0, \ldots, h_d\}$, $\sum_{j=1}^{d} \theta_j = 1$ and:

$$C(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{S}} \prod_{j=1}^{d} \binom{h_j}{x_j} \theta_x^{x_j} \tag{7.8}$$

The decomposition (7.2) is obtained with

$$\mathcal{L}(\boldsymbol{\theta}_j) = \mathcal{B}\left(h_j, \frac{t\theta_j}{1 + t\theta_j}\right)$$

the binomial distribution with mean $h_j \left(\frac{t\theta_j}{1+t\theta_j}\right)$ for any $t \in (0, +\infty)$, and Y_t has no closed-form distribution.

Definition 7.5. The multi-dimensional non-central Pólya distribution $\mathcal{P}_d^*(N, \boldsymbol{q}, \boldsymbol{\theta})$ is defined by:

$$\begin{aligned} &\forall \boldsymbol{x} \in \mathbb{N}^d, \\ \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) &= \left(\frac{\prod_{j=1}^{d} \binom{x_j + q_j - 1}{x_j} \theta_j^{x_j}}{D(\boldsymbol{\theta})} \right) \mathbb{1}_{\mathcal{S}}(\boldsymbol{x}) \end{aligned} \tag{7.9}$$

where $\forall j \in \{1, \ldots, d\}, q_j > 0$, $\sum_{j=1}^{d} \theta_j = 1$ and:

$$D(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{S}} \prod_{j=1}^{d} \binom{x_j + q_j - 1}{x_j} \theta_j^{x_j}$$

The decomposition (7.2) is obtained with

$$\mathcal{L}(\boldsymbol{\theta}_j) = \mathcal{NB}(q_j, t\theta_j)$$

the negative binomial distribution with mean $q_j \left(\frac{1-t\theta_j}{t\theta_j}\right)$ for any $t \in (0, 1)$, and Y_t has no closed-form distribution.

To contrast the performances of our algorithm with respect to the exact ones, let us introduce the following notations:

$$\sigma_a = \sum_{j=1}^{d} a_j, \quad \sigma_b = \sum_{j=1}^{d} b_j, \quad \sigma_{ab} = \sigma_b - \sigma_a, \quad N_a = N - \sigma_a \tag{7.10}$$

We note that if $N < \sigma_a$ or $\sigma_b > N$ then $p_{\mathcal{D}}(\boldsymbol{a}, \boldsymbol{b}) = 0$. We also note that if $N = \sigma_a$ or $N = \sigma_b$, the evaluation of $p_{\mathcal{D}}(\boldsymbol{a}, \boldsymbol{b})$ reduces to the evaluation of the probability function at $\boldsymbol{a}$ or $\boldsymbol{b}$, and if $\sigma_{ab} = 0$ with $\boldsymbol{a} \leq \boldsymbol{b}$, then $\boldsymbol{a} = \boldsymbol{b}$ and it reduces also to the evaluation of the probability function at $\boldsymbol{a}$. The only difficult case is then when $\sigma_a < N < \sigma_b$.

For all the distributions $\mathcal{L}(\boldsymbol{\theta})$ we are interested in, there exist efficient and accurate routines to evaluate both $\mathbb{P}(a_j \leq Y_{tj} \leq b_j)$ and $\mathbb{P}(Y_t = N)$. The only difficulty is the evaluation of $\mathbb{P}(T_t = N)$, as noticed in [Goo57], [Lev81], [Lev83], [Lev92] and [Ma99], but they gave no clue on how to do it both efficiently and accurately. Instead, they developed several approximations of this quantity using either Edgeworth expansions or saddle-point approximations. The approximate algorithm proposed in [BS98] results from the following second order saddlepoint approximation of $\mathbb{P}(T_t = N)$:

$$\mathbb{P}(T_t = N) \simeq \frac{e^{K_t(\hat{s}) - N\hat{s}}}{\sqrt{2\pi K_t''(\hat{s})}} \left(1 + \frac{1}{8}\frac{K_t^{(4)}(\hat{s})}{K_t''(\hat{s})^2} + \frac{5}{24}\frac{K_t^{(3)}(\hat{s})^2}{K_t''(\hat{s})^3}\right)$$

where $K_t(s)$ is the cumulant generating function $\log \mathbb{E}\left[e^{sT_t}\right]$ of T_t and the saddlepoint $\hat{s}$ is the unique solution to $K_t'(s) = N$, which is guaranteed to exist if $\sigma_a < N < \sigma_b$. As $T_t = \sum_{j=1}^{d} T_{tj}$ where the random variable T_{tj} are independent, we have $K_t(s) = \sum_{j=1}^{d} K_{tj}(s)$.

For the case of the evaluation of multinomial rectangular probabilities, $K_{tj}(s)$ reads:

$$\forall s \in \mathbb{R},\ K_{tj}(s) = \log\left(\sum_{k=a_j}^{b_j} e^{ks - tp_j}\frac{(tp_j)^k}{k!}\right) - \log\left(\sum_{k=a_j}^{b_j} e^{-tp_j}\frac{(tp_j)^k}{k!}\right)$$

and the saddlepoint $\hat{s}$ has to be found numerically. The approximate algorithm proposed in [Lev81] results from the following second-order Edgeworth expansion of $\mathbb{P}(T_t = N)$:

$$\mathbb{P}(T_t = N) \simeq \frac{e^{-z^2/2}}{\sqrt{2\pi K_t''(0)}}\left(1 + \frac{1}{6}\frac{K_t^{(3)}(0)}{K_t''(0)^{3/2}}H_3(z) + \frac{1}{24}\frac{K_t^{(4)}(0)}{K_t''(0)^2}H_4(z) + \frac{1}{72}\frac{K_t^{(3)}(0)^2}{K_t''(0)^3}H_6(z)\right)$$

where $z = \frac{N - K_t'(0)}{\sqrt{K_t''(0)}}$ is the standardized value of N and H_3, H_4 and H_6 are the third, fourth and sixth degree Hermite polynomials.

While both approximations involve the parameter t, in the case of the saddlepoint approximation this parameter cancel out with the remaining of (7.2) when $\mathbb{P}(T_t = N)$ is replaced by its approximation. On the contrary, t does not cancel out when the approximation based on Edgeworth expansion is used in (7.2), thus a choice has to be made for t in that case, and the value of $t = N$ is recommanded in [Lev81].

The space complexity of both approximations is $\mathcal{O}(1)$ as $\sigma_{ab} \to \infty$ and the time complexity is $\mathcal{O}\left(\sigma_{ab}\log_2\left(\frac{\sigma_{ab}}{\epsilon}\right)\right)$ as $\sigma_{ab} \to \infty, \epsilon \to 0$ for the saddlepoint approximation if $\hat{s}$ is computed using bisection on the interval $[\sigma_a, \sigma_b]$ with a precision of ϵ, and $\mathcal{O}(N)$ as $N \to \infty$ for the Edgeworth expansion approximation.

Reference	[BM95]	[Cor11]	[Fre09]
Distributions	$\mathcal{M}_d$	$\mathcal{M}_d$, $\mathcal{H}_d$	$\mathcal{M}_d$
Restrictions on $\boldsymbol{a}$, $\boldsymbol{b}$	$\boldsymbol{a} = \mathbf{0}$	$\boldsymbol{a} = (a, \dots, a)$ $\boldsymbol{b} = (b, \dots, b)$	**none**

Table 7.1: Exact algorithms applicability

Reference	[BS98],[Lev83], [Lev92]	[CB00]	[Goo57]
Distributions	$\mathcal{M}_d$, $\mathcal{H}_d$, $\mathcal{P}_d$	$\mathcal{H}_d$	$\mathcal{M}_d$
Restrictions on $\boldsymbol{a}$, $\boldsymbol{b}$	**none**	**none**	$\boldsymbol{a} = (a, \dots, a)$ $\boldsymbol{b} = (b, \dots, b)$
Reference	[Ma99]	[Lev81]	
Distributions	$\mathcal{H}_d^*$, $\mathcal{P}_d^*$	$\mathcal{M}_d$	
Restrictions on $\boldsymbol{a}$, $\boldsymbol{b}$	$\boldsymbol{a} = \mathbf{0}$, $\boldsymbol{b} = (b, \dots, b)$ or $\boldsymbol{a} = (a, \dots, a)$, $\boldsymbol{b} = (\infty, \dots, \infty)$	$\boldsymbol{a} = \mathbf{0}$	

Table 7.2: Approximate algorithms applicability

According to the theoretical analysis and numerical study presented in [BS98], the saddlepoint approximation is the most accurate approximation, with a measured relative error ranging from 0.005% to 5% on a set of configurations for multinomial rectangular probabilities covering dimensions d from 8 up to 200 and population size N from 12 to 300, while the Edgeworth-based approximation gives relative errors ranging from 0.01% to 5% on the same configurations.

Is the accurate (or even exact) evaluation $\mathbb{P}(T_t = N)$ intractable? A straightforward approach is to compute the associated convolution by multiplying the generating functions of the Y_{tj} random variables and by extracting the coefficient of degree N. All these generating functions are polynomials of degree b_j, of which only the coefficients of order not greater than N are of interest. It leads to a $\mathcal{O}(dN^2)$ time complexity as $d, N \to \infty$ if these multiplications are done using the naive polynomial multiplication algorithm, or to a $\mathcal{O}(dN \log N)$ time complexity as $d, N \to \infty$ if these multiplications are done using a FFT based algorithm. While it is clearly a much better algorithm than the brute force enumeration method, it remains costly for large values of N and d.

The most efficient exact algorithm proposed so far for the evaluation of (7.1), in the case of the multinomial distribution, is the one described in [Fre09] and recalled below. Its space complexity is $\mathcal{O}(\sigma_{ab})$ as $\sigma_{ab} \to \infty$ and its time complexity is $\mathcal{O}(N_a \sigma_{ab})$ as $N_a, \sigma_{ab} \to \infty$. The algorithm described in [Cor11] has the same space and time complexity, covers more distributions but for restricted arguments of (7.1), so we take [Fre09] as a reference in terms of accuracy, space and time complexity. These algorithms are not based on the representation (7.2), but rather on an incremental evaluation of the summation of the probability function associated to the evaluation of rectangular probabilities. We detail here the algorithm presented in [Fre09] in the next section.

Using an appropriate numerical method, it is possible to evaluate the order N coeffi-

cient efficiently in both space and time, not exactly but with a user-controlled accuracy that can be made as small as the machine precision. In some sense, the resulting algorithm is essentially exact.

To jump directly to the conclusion, and to motivate the reader, the key results concerning the new algorithm we propose are that it has a constant (and small) $\mathcal{O}(1)$ space complexity as $\sigma_{ab} \to \infty$, and has a worst case $\mathcal{O}(N_a \sigma_{ab})$ time complexity as $N_a, \sigma_{ab} \to \infty$ that drops to $\mathcal{O}(d\sqrt{N_a})$ as $N_a, \sigma_{ab} \to \infty$ for most situations, for a relative precision comparable to [Fre09], which is a tremendous improvement with respect to the best current exact algorithms.

The first section of the chapter presents the reference algorithm proposed in [Fre09]. The second section presents the foundations of the new algorithm and the third section details some specific results that make the algorithm efficient, with a particular emphasize on the multinomial case. The last section gives experimental evidences of both the time complexity and the accuracy of the algorithm. Several test cases gathered in the literature are also detailed.

7.2 Reference algorithm

We present here the algorithm proposed in [Fre09], which is an exact algorithm computing multinomial rectangular probabilities and which is based on an incremental evaluation of the probabilistic contribution of all the integer-valued coordinates points in the rectangular region of interest. We focus on the case where $\sigma_a < N < \sigma_b$, as the other cases lead either to a null probability or to a contribution limited to one point, which is readily evaluated using (7.4).

The multinomial rectangular probability $p_{\mathcal{M}}(\boldsymbol{a}, \boldsymbol{b})$ can be writtten as follows:

$$\begin{aligned} p_{\mathcal{M}}(\boldsymbol{a}, \boldsymbol{b}) &= \sum_{x_1=a_1}^{b_1} \cdots \sum_{x_d=a_d}^{b_d} \left(\frac{N!}{x_1! \cdots x_d!} \right) p_1^{x_1} \cdots p_d^{x_d} \mathbb{1}_{\mathcal{S}}(\boldsymbol{x}) \\ &= \left(\frac{N! p_1^{a_1} \cdots p_d^{a_d}}{a_1! \cdots a_d!} \right) \sum_{x_1=a_1}^{b_1} \cdots \sum_{x_d=a_d}^{b_d} \left\{ \prod_{i=1}^{d} \prod_{j=1}^{x_i - a_i} \left(\frac{p_i}{a_i + j} \right) \right\} \mathbb{1}_{\mathcal{S}}(\boldsymbol{x}) \end{aligned} \tag{7.11}$$

where the inner product is understood to be 1 if $x_i - a_i = 0$. Each non-zero summand is the product of N_a factors that can be associated in a one-to-one fashion to a vector $\boldsymbol{\ell}$ of indices $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_{N_a})$ such that the first $(x_1 - a_1)$ entries are 1, the next $(x_2 - a_2)$ entries are 2, and so one. The first $(x_1 - a_1)$ entries are for the terms of the form $\frac{p_1}{a_1 + j}$, the next $(x_2 - a_2)$ entries are for the terms of the form $\frac{p_2}{a_2 + j}$ and so on.

The set K of non-zero summands is then the set of all nondecreasing vectors $\boldsymbol{\ell}$ in $\{1, \ldots, d\}^{N_a}$ such that each value j appears no more than $b_j - a_j$ times. We can then write $p_{\mathcal{M}}(\boldsymbol{a}, \boldsymbol{b})$ as:

$$p_{\mathcal{M}}(\boldsymbol{a}, \boldsymbol{b}) = \left(\frac{N! p_1^{a_1} \cdots p_d^{a_d}}{a_1! \cdots a_d!} \right) \sum_{\boldsymbol{\ell} \in K} F_1(\ell_1) F_2(\ell_1, \ell_2) \cdots F_{N_a}(\ell_1, \ldots, \ell_{N_a}) \tag{7.12}$$

where $F_1(\ell_1)$,...,$F_{N_a}(\ell_1, \ldots, \ell_{N_a})$ are the factors $\frac{p_{\ell_1}}{a_{\ell_1} + j}$ that contribute to non-zero summands in (7.11).

The key point of the algorithm is to evaluate (7.12) in a recursive way. To this end, we introduce the set $K_m(i, j)$ defined as the set of all nondecreasing vectors of integers $\boldsymbol{\ell} \in \{1, \ldots, d\}^m$ so that:

1. $\ell_{m-j} < \ell_{m-j+1} = \cdots = \ell_m = i$
2. $\boldsymbol{\ell}$ does not contain any value i more than $b_i - a_i$ times.

By defining $P_m(i,j)$ by:

$$P_m(i,j) = \sum_{\ell \in K_m(i,j)} F_1(\ell_1) F_2(\ell_1,\ell_2) \cdots F_m(\ell_1,\ldots,\ell_m)$$

for all $m \in \{1,\ldots,N_a\}$, $i \in \{1,\ldots,d\}$ and $j \in \{1,\ldots,b_i - a_i\}$, the multinomial rectangular probability $p_{\mathcal{M}}(\boldsymbol{a},\boldsymbol{b})$ can be expressed as:

$$p_{\mathcal{M}}(\boldsymbol{a},\boldsymbol{b}) = \left(\frac{N! p_1^{a_1} \cdots p_d^{a_d}}{a_1! \cdots a_d!}\right) \sum_{i=1}^{d} \sum_{j=1}^{x_i - a_i} P_d(i,j) \tag{7.13}$$

The values $P_m(i,j)$ are evaluated recursively thanks to:

$$P_1(i,j) = \begin{cases} \dfrac{p_i}{a_i + 1} & j = 1 \text{ and } b_i > a_i \\ 0 & \text{otherwise} \end{cases} \tag{7.14}$$

for the initialization and to

$$P_{m+1}(i,j) = \begin{cases} \left(\dfrac{p_i}{a_i+1}\right) \displaystyle\sum_{k=1}^{i-1} \sum_{j=1}^{\min(b_i - a_i, m)} P_m(k,j) & \text{for } j = 1 \\ \left(\dfrac{p_i}{a_i + j}\right) P_m(i, j-1) & \text{for } j > 1 \end{cases} \tag{7.15}$$

for the recursion.

The resulting algorithm is then:

Algorithm 7.6.

Given $\boldsymbol{a},\boldsymbol{b} \in \mathbb{N}^d$ such that $\boldsymbol{a} \leq \boldsymbol{b}$, $N \in \mathbb{N}$ such that $\sigma_a < N < \sigma_b$ and positive probabilities $p_1,\ldots,p_d$ such that $\sum_{j=1}^{d} p_j = 1$, do

1. Compute the N_a values of $P_1(i,j) : i = 1,\ldots,d;\ j = 1,\ldots,b_i - a_i$ using (7.14)
2. For $m = 1,\ldots,N_a - 1$, compute the N_a values of $P_{m+1}(i,j) : i = 1,\ldots,d;\ j = 1,\ldots,b_i - a_i$ from $P_m(i,j) : i = 1,\ldots,d;\ j = 1,\ldots,b_i - a_i$ using (7.15).
3. Compute $p_{\mathcal{M}}(\boldsymbol{a},\boldsymbol{b})$ using (7.13).

The space complexity of this algorithm is $\mathcal{O}(\sigma_{ab})$ as $\sigma_{ab} \to \infty$ due to the storage of the N_a values of $P_m(i,j) : i = 1,\ldots,d;\ j = 1,\ldots,b_i - a_i$ and its time complexity is $\mathcal{O}(N_a \sigma_{ab})$ as $N_a, \sigma_{ab} \to \infty$ because m runs from 1 to N_a. The final step has a cost of $\mathcal{O}(N_a)$ as $N_a \to \infty$ and is thus negligible with respect to the other steps.

7.3 Foundations of the new algorithm

In this section, our key result is the representation of the rectangular probability given in Proposition 7.11, which is the basis of our new algorithm.

Here is the key result given in [AW91, equations 5.35–5.37] and allowing for a fast and accurate evaluation of convolutions for discrete univariate distributions, using the Poisson summation formula:

Theorem 7.7. *Let f_X be the probability function of a discrete random variable X and ϕ_X its associated probability generating function $\phi_X(z) = \sum_{k\geq 0} f_X(k) z^k$.*

Then, for any non-negative integers n, $m > n$ and real number $0 < r < 1$:

$$f_X(n) = \mathbb{P}(X = n) = \frac{1}{mr^n} \sum_{k=0}^{m-1} \xi_m^{-kn} \phi_X\left(r\xi_m^k\right) - \epsilon_{n,m,r} \tag{7.16}$$

where $\epsilon_{n,m,r} = \sum_{k\geq 1} f_X(n+km) r^{km} \leq \mathbb{P}(X \geq m+n) r^m \leq r^m$ and $\xi_m = e^{\frac{2i\pi}{m}}$.

We recall the proof of this result for the reader convenience.

Proof. The generating function ϕ_X is defined at least for $z \in \mathbb{C}$ such that $|z| < 1$. For such a z, using polar coordinates we have:

$$\forall r \in (0,1), \theta \in [0, 2\pi), \quad \phi_X\left(re^{i\theta}\right) = \sum_{j\geq 0} f_X(j) r^j e^{ij\theta} = \sum_{j\in\mathbb{Z}} a_j e^{ij\theta}$$

where the sequence of real numbers $(a_j)_{j\in\mathbb{Z}}$ is defined by:

$$\forall j \in \mathbb{Z}, \quad a_j = \begin{cases} 0 & \text{if } j < 0 \\ f_X(j) r^j & \text{if } j \geq 0 \end{cases}$$

Let $m \in \mathbb{N}^*$ be a positive integer. We form the m-periodic sequence $\left(a_j^m\right)_{j\in\mathbb{Z}}$ defined by:

$$\forall j \in \{0, \ldots, m-1\}, \quad a_j^m = \sum_{\ell\in\mathbb{Z}} a_{j+\ell m}$$

Its discrete Fourier transform $\left(b_k^p\right)_{k\in\{0,\ldots,m-1\}}$ reads:

$$\begin{aligned} \forall k \in \{0, \ldots, m-1\}, \quad b_k^m =& \frac{1}{m} \sum_{j=0}^{m-1} a_j^m e^{2i\pi jk/m} \\ =& \frac{1}{m} \sum_{j=0}^{m-1} \sum_{\ell\in\mathbb{Z}} a_{j+\ell m} e^{2i\pi jk/m} \\ =& \frac{1}{m} \sum_{j\in\mathbb{Z}} a_j e^{2i\pi jk/m} \\ =& \frac{1}{m} \phi_X\left(r\xi_m^k\right) \end{aligned}$$

where $\xi_m = e^{\frac{2i\pi}{m}}$. Using the inversion formula of the discrete Fourier transform, we get:

$$\begin{aligned} \forall j \in \{0, \ldots, m-1\}, \quad a_j^m =& \sum_{k=0}^{m-1} b_k^m e^{-2i\pi jk/m} \\ =& \frac{1}{m} \sum_{k=0}^{m-1} \xi_m^{-jk} \phi_X\left(r\xi_m^k\right) \end{aligned}$$

from which we deduce that $f_X(n) = \frac{1}{mr^n} \sum_{k=0}^{m-1} \xi_m^{-kn} \phi_X\left(r\xi_m^k\right) - \epsilon_{n,m,r}$ with $n \in \mathbb{N}$, $m \in \mathbb{N}$ such that $m > n$, $r \in (0,1)$ and:

$$\begin{aligned} \epsilon_{n,m,r} =& \sum_{k\geq 1} f_X(n+km) r^{km} \\ \leq& \mathbb{P}(X \geq m+n) r^m \\ \leq& r^m \end{aligned}$$

as $\forall k \geq 1$, $r^{km} < r^m$ and $\sum_{k\geq 1} f_X(n+km) \leq \sum_{j\geq 0} f_X(n+m+j)$. □

This theorem provides a numerical method to compute the probability function of a discrete distribution from its generating function: the value of $f_X(n)$ is approximated by the finite sum that appears in (7.16), with a positive error (i.e. $f_X(n)$ is **over-estimated**) that can be made as small as needed by a judicious choice of r and m. We note that if X has a bounded support with upper bound M, which is the case in the application we have in mind, any choice of m such that $m > M$ leads to an **exact** algorithm as $\epsilon_{n,m,r} = 0$ for such a choice.

Using $m = 2n$ in (7.16) gives two advantages, namely the terms of the sum can be paired in order to add to real values so the resulting formula has no more than $n + 1$ terms, and the factor ξ_m^{-kn} reduces to $(-1)^k$. The resulting formula is given in [AW91, equations 5.38–5.39], and reads:

Proposition 7.8. *If we take $m = 2n$ in (7.16), we get:*

$$f_X(n) = \frac{1}{2nr^n} \sum_{k=0}^{n-1} (-1)^k \Re\left(\phi_X\left(r\zeta_n^k\right) - \phi_X\left(r\zeta_n^{k+1}\right)\right) - \epsilon_{n,r} \tag{7.17}$$

where

$$\epsilon_{n,r} = \sum_{k\geq 1} f_X((2k+1)n) r^{2kn} \leq \mathbb{P}(X \geq 3n) r^{2n} \leq r^{2n} \tag{7.18}$$

and $\zeta_n = \xi_{2n} = e^{\frac{i\pi}{n}}$

We will apply (7.16) if we want an exact algorithm, or (7.17) if we want an approximate algorithm, to evaluate $\mathbb{P}(T_t = N)$, and plug the resulting formula into (7.2) in order to derive our algorithm. In the approximate case, we see that the value of the error (7.18) can be made smaller than a given ϵ_{max} by choosing r such that $r^{2n} \leq \epsilon_{max}$:

$$r \leq \epsilon_{max}^{\frac{1}{2n}} \tag{7.19}$$

It remains to express the generating probability function of T_t, which is an elementary result stated without proof:

Proposition 7.9. *$\forall j \in \{1, \ldots, d\}$, $\forall z \in \mathbb{C}$ with $|z| \leq 1$ we have:*

$$\phi_{T_{tj}}(z) = \frac{\pi_{a_j b_j}^{(j)}(z)}{\mathbb{P}(a_j \leq Y_{tj} \leq b_j)} \tag{7.20}$$

with

$$\pi_{a_j b_j}^{(j)}(z) = \sum_{k=a_j}^{b_j} \mathbb{P}(Y_{tj} = k) z^k = \pi_{b_j}^{(j)}(z) - \pi_{a_j - 1}^{(j)}(z) \tag{7.21}$$

where

$$\pi_{-1}^{(j)}(z) \equiv 0, \quad \forall n \in \mathbb{N}, \pi_n^{(j)}(z) = \sum_{k=0}^{n} \mathbb{P}(Y_{tj} = k) z^k \tag{7.22}$$

The independence of the T_{tj} leads to:

$$\phi_{T_t}(z) = \frac{\prod_{j=1}^d \pi_{a_j b_j}^{(j)}(z)}{\prod_{j=1}^d \mathbb{P}(a_j \leq Y_{tj} \leq b_j)} \tag{7.23}$$

We see that a key factor in the cost of (7.16) is the constraint $m > n$. When we are interested in computing the value of a multi-dimensional discrete distribution function, there is no choice but to take $n = N$ and $m > N$ in (7.16). But when we are interested in computing a rectangular probability, i.e. when $\boldsymbol{a} \neq \boldsymbol{0}$, we can express $\mathbb{P}(T_t = N)$ in a form that leads to a less expensive summation. The elementary properties of the characteristic functions lead to the following proposition:

Proposition 7.10. *If $\boldsymbol{a} \neq \boldsymbol{0}$, $\forall j \in \{1, \dots, d\}$ we set $V_{tj} = T_{tj} - a_j$. The random variables V_{tj} are such that:*

$$\mathbb{P}(V_{tj} = k) = \mathbb{P}(T_{tj} = a_j + k) \tag{7.24}$$

$$\phi_{V_{tj}}(z) = z^{-a_j} \phi_{T_{tj}}(z) \tag{7.25}$$

and

$$\mathbb{P}(T_t = N) = \mathbb{P}(V_t = N_a) \tag{7.26}$$

$$\phi_{V_t}(z) = z^{-\sigma_a} \phi_{T_t}(z) \tag{7.27}$$

where $V_t = \sum_{j=1}^{d} V_{tj}$ has support $\{0, \dots, \sigma_{ab}\}$.

Replacing the evaluation of $\mathbb{P}(T_t = N)$ by the evaluation of $\mathbb{P}(V_t = N_a)$ moves the constraint $m > N$ into $m > N_a$ with $N_a < N$. Furthermore, the algorithm is now exact as soon as $m > \sigma_{ab}$.

Considering only the approximate version of the algorithm, we get:

Proposition 7.11. *$\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{N}^d$, we have:*

$$\begin{aligned} p_{\mathcal{D}}(\boldsymbol{a}, \boldsymbol{b}) \quad = \quad & \Re\left\{ \sum_{k=0}^{N-1} (-\zeta_{N_a}^{-\sigma_a})^k \left(\prod_{j=1}^{d} \pi_{a_j b_j}^{(j)}\left(r\zeta_{N_a}^{k}\right) - \zeta_{N_a}^{-\sigma_a} \prod_{j=1}^{d} \pi_{a_j b_j}^{(j)}\left(r\zeta_{N_a}^{k+1}\right) \right) \right\} / \\ & \left(2N_a r^N \mathbb{P}(Y_t = N)\right) - \eta_{N_a, r} \end{aligned} \tag{7.28}$$

where $K = \dfrac{\prod_{j=1}^{d} \mathbb{P}(a_j \leq Y_{tj} \leq b_j)}{\mathbb{P}(Y_t = N)}$ and $\eta_{N_a,r} = K\epsilon_{N_a,r}$.

Except for the storage of the data $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{p}$, which is a $\mathcal{O}(d)$ as $d \to \infty$, the memory complexity of this algorithm is $\mathcal{O}(1)$ as $N, d \to \infty$ because no intermediate structure is needed in the evaluation of (7.28). The time complexity is of order $\mathcal{O}(N_a C)$ as $N_a, C \to \infty$, where C is the time complexity of evaluating $\pi_{a_1 b_1}^{(j)}, \dots, \pi_{a_d b_d}^{(j)}$ at a given point. A naive evaluation of these polynomials leads to $C \simeq \sigma_{ab}$ and a total time complexity of $\mathcal{O}(N_a \sigma_{ab})$ as $N_a, \sigma_{ab} \to \infty$, which is the same complexity as the algorithm proposed in [Fre09]. We also note that the factor $\prod_{j=1}^{d} \mathbb{P}(a_j \leq Y_{tj} \leq b_j)$ in (7.1) simplifies with the denominator of (7.23), reducing the overall computational cost.

One can see that the error in (7.28) depends on t through the numerator of K, and on r through $\epsilon_{N_a,r}$. The theoretical behavior of this error is clear: we can take r small enough to get the absolute error we want. The numerical behavior of this error is less clear as the summation in (7.28) can be subject to cancellation, increasing the error. The best way to take into account these cancellations is to use the recommendations in [AW91] to choose r using (7.19), then to choose t in order to minimize K. This point will be explored numerically in the case of the multinomial distribution.

7.4 Making the new algorithm more efficient

In this section, our key results are the efficient evaluation of the characteristic function of T_t, as a result of Proposition 7.12, and the original stopping criterion given in Proposition 7.14. Combined, these results lead to Algorithms 7.13 and 7.15, which are our core contribution.

Two remarks can lead to a dramatic improvement of the time complexity (or complexity for short) of the proposed algorithm. The first one is that in many situations, the evaluation of $\pi^{(j)}_{a_j b_j}$ can be done with $\mathcal{O}(1)$ operations instead of $\mathcal{O}(b_j - a_j)$ within machine precision when $b_j - a_j \to \infty$, counting the evaluation of a transcendental function such as exp as an unitary cost operation. In this case, $C = \mathcal{O}(d)$ as $d \to \infty$ instead of $C = \mathcal{O}(\sigma_{ab})$ as $\sigma_{ab} \to \infty$, and the total complexity drops to $\mathcal{O}(N_a d)$ as $N_a, d \to \infty$. The second one is that the terms involved in (7.28) are usually of very different magnitudes, and most of them do not contribute significantly (up to machine precision) to the final result. It is common that only $\mathcal{O}(\sqrt{N_a})$ terms are needed as $N_a \to \infty$. The overall complexity is thus reduced to $\mathcal{O}(d\sqrt{N_a})$ as $N_a, d \to \infty$.

7.4.1 Efficient evaluation of $\pi^{(j)}_{a_j,b_j}(z)$

If $b_j - a_j = \mathcal{O}(1)$ as $b_j - a_j \to \infty$, the evaluation of $\pi^{(j)}_{a_j b_j}$ is obviously a $\mathcal{O}(1)$ as $b_j - a_j \to \infty$, so we restrict our attention to the case $b_j - a_j \to \infty$. It covers two different sub-cases: either we have a_j of order the unity, for example in the case where one is interested in the computation of the distribution function of the distribution, or we have $a_j, b_j \to \infty$. In the first case, the following proposition gives elements to make the evaluation of $\pi^{(j)}_{a_j b_j}$ cheaper than $\mathcal{O}(b_j - a_j)$ as $b_j - a_j \to \infty$:

Proposition 7.12. *Let n be a nonnegative integer and z a complex number such that $|z| \leq 1$. Let $\bar{s} = \sup\{s \geq 0, \phi_{Y_{tj}}(e^s) < +\infty\}$. If $\bar{s} > 0$, then*

$$\left|\phi_{Y_{tj}}(z) - \pi^{(j)}_n(z)\right| \leq \frac{\phi_{Y_{tj}}\left(e^{s^*}\right)}{e^{(n+1)s^*}} \tag{7.29}$$

where $s^* = \underset{0<s<\bar{s}}{\operatorname{argmin}} \dfrac{\phi_{Y_{tj}}(e^s)}{e^{(n+1)s}}$.

Proof. By definition of $\phi_{Y_{tj}}$ and $\pi^{(j)}_n(z)$, we have:

$$\begin{aligned}
\left|\phi_{Y_{tj}}(z) - \pi^{(j)}_n(z)\right| &= \left|\sum_{k>n} \mathbb{P}\left(Y_{tj} = k\right) z^k\right| \\
&\leq \sum_{k>n} \mathbb{P}\left(Y_{tj} = k\right) |z|^k \\
&\leq F^c_{Y_{tj}}(n) \text{ as } |z| \leq 1
\end{aligned}$$

where $F^c_{Y_{tj}}(n) = \mathbb{P}\left(Y_{tj} > n\right)$ is the complementary distribution function of Y_{tj} evaluated at n.

Then, applying the Markov inequality to $e^{sY_{tj}}$ and minimizing the bound with respect to s such that $0 < s < \bar{s}$ we get (7.29). □

The hypothesis made on $\phi_{Y_{tj}}$ is fulfilled in the particular cases of the Poisson, binomial and negative binomial distributions, for which the respective values of s^* are $\log\left(\frac{n+1}{tp_j}\right)$,

$\log\left(\frac{1-t}{t}\frac{n+1}{h_j-(n+1)}\right)$ and $\log\left(\frac{n+1}{(1-t)(q_j+n+1)}\right)$. In the general case, the convergence of $\pi_n^{(j)}(z)$ to $\phi_{Y_{tj}}(z)$ is at least exponential with n, and can be even faster in specific cases (e.g. in the Poisson case). It results the following algorithm to evaluate $\pi_n^{(j)}(z)$:

Algorithm 7.13.

Given $n \in \mathbb{N}$, $z \in \mathbb{C}$, $|z| \leq 1$, do:

1. Set $k := n+1$
2. Set $v_k := \phi_{Y_{tj}}(z)$
3. Set $dv_k = \mathbb{P}(Y_{tj} = k)z^k$
4. While $|dv_k| > |v_k|\epsilon_{machine}$ do
   (a) $v_{k+1} := v_k - dv_k$
   (b) $dv_{k+1} := \mathbb{P}(Y_{tj} = k+1)z^{k+1} = f(dv_k, k, z)$
   (c) $k := k+1$
5. Return v_k

This algorithm performs $\mathcal{O}(|\log \epsilon_{machine}|)$ iterations as $\epsilon_{machine} \to 0$. The evaluation of $\phi_{Y_{tj}}(z)$ and $\mathbb{P}(Y_{tj} = k)$ can be done in constant time complexity for the common distributions, and the update (4b) can be made for usual distributions using a simple recursion $f(dv_k, k, z)$ instead of the full evaluation of $\mathbb{P}(Y_{tj} = k+1)z^{k+1}$.

Considering the case of the multinomial distribution, i.e. $Y_{tj} \sim \mathcal{P}(tp_j)$, the situation of (7.12) is likely to occur when $\sigma_b = \mathcal{O}(dN)$ as $N, d \to \infty$ and $\sigma_a = \mathcal{O}(d)$, i.e. when $b_j = \mathcal{O}(N)$ and $a_j = \mathcal{O}(1)$, which corresponds to the worst complexity we get using the naive evaluation of all the $\pi_{a_j b_j}^{(j)}$. In this case, the number of iterations of (7.13) is less than 18 for $\epsilon_{machine} = 10^{-16}$ and $tp_j = 1$, a situation typical of interacting particle algorithms setting. The update is given by $f(dv_k, k, z) = dv_k \times \frac{tp_j z}{k+1}$.

When both a_j and b_j are large, the situation is more involved. A naive evaluation using $\pi_{a_j,b_j}^{(j)} = \pi_{b_j}^{(j)} - \pi_{a_j-1}^{(j)}$ can suffer from massive cancellation, providing a very inaccurate result. Nevertheless, in the multinomial case, a systematic $\mathcal{O}(1)$ time complexity can be achieved for the evaluation of $\pi_{a_j,b_j}^{(j)}(z)$, in connection with the evaluation of the regularized incomplete gamma function, see [DM86], [Tem94] and the boost library (www.boost.org) for an efficient implementation of these methods.

7.4.2 Fast (essentially) exact evaluation of the Poisson summation

The terms involved in (7.28) can have very different magnitudes. As a result, only a few of them could have a significant contribution to the Poisson summation formula, and taking advantage of it could reduce very significantly the cost in the evaluation of the sum. We illustrate it in the computation of the multinomial distribution function. In this case, $N_a = N$, $V_{tj} = T_{tj}$ and $V_t = T_t$. We restrict our analysis to the case where $\phi_{T_t} \simeq \phi_{Y_t} = e^{-t(1-z)}$:

Proposition 7.14. *We consider the case where $N \gg 1$ and $\phi_{T_t}(z) \simeq e^{-t(1-z)}$. Either $t = \mathcal{O}(1)$ and no term of the sum in (7.28) is negligible, or $t \to +\infty$ with $t = \mathcal{O}(N)$ and only the N^* first terms of (7.28) have a relative contribution to $p_{\mathcal{D}}(\boldsymbol{a}, \boldsymbol{b})$ greater than $\epsilon \ll 1$, with:*

$$N^* \simeq \frac{1}{\pi}\sqrt{-\frac{2\log\epsilon}{r}\frac{N^2}{t}} \tag{7.30}$$

The case $t \gg N$ is not relevant, as it leads to severe cancellations in (7.28).

Proof. In order to study the magnitude of the terms occurring in (7.17), we use the elementary relation:

$$|e^{\beta} - e^{\alpha}|^2 = e^{2\Re(\beta)}\rho_{\alpha,\beta} \tag{7.31}$$

with

$$\rho_{\alpha,\beta} = 1 + e^{2\Re(\alpha-\beta)} - 2\cos\left(\Im(\alpha-\beta)\right)e^{\Re(\alpha-\beta)} \tag{7.32}$$

using $\beta = -t\left(1 - r\zeta_N^k\right)$ and $\alpha = -t\left(1 - r\zeta_N^{k+1}\right)$.

Let ρ_k be the value of $\rho_{\alpha,\beta}$ for this choice of α and β, and $\delta_k = \phi_{Tt}\left(r\zeta_N^k\right) - \phi_{Tt}\left(r\zeta_N^{k+1}\right)$. Three cases have to be considered: $t = \mathcal{O}(1)$, $t = o(N)$ with $t \to +\infty$ and $t = \Theta(N)$ [1]. For the first and second cases, using $\zeta_N - 1 = \frac{i\pi}{N} + \mathcal{O}\left(\frac{1}{N}\right)$ we get:

$$\frac{|\delta_k|}{|\delta_0|} = e^{-rt(1-\cos(\theta_k))} + \mathcal{O}\left(\frac{t}{N}\right) \tag{7.33}$$

We have $|\delta_k|/|\delta_0| < \epsilon$ as soon as $\cos(\theta_k) \leq 1 + \frac{\log\epsilon}{rt}$. For typical values of ϵ, when $t = \mathcal{O}(1)$, it is not possible to fulfill this constraint so one must compute the N terms in (7.17). In the second case, using the expansion $\arccos(1-x) = \sqrt{2x} + \mathcal{O}(x^{3/2})$ we get the value of N^* given in (7.30).

When $t = \Theta(N)$, the computation is more involved as the terms in t/N are no more negligible. We proceed in two steps: first we show that $\frac{|\delta_k|}{|\delta_0|}$ is small as soon as k is greater than a bound which is a $o(N)$, justifying that one can use series expansions with respect to $\theta_k = \frac{k\pi}{N} = o(1)$, then one gets (7.30) by computations similar to the previous cases. More precisely, defining $\gamma = \frac{t}{N}$ and assuming that $\cos(\gamma r\pi) < 1$, we get:

$$\begin{aligned}\frac{|\delta_k|}{|\delta_0|} &= e^{-rt(1-\cos(\theta_k))}\sqrt{\frac{1 - 2\cos(\gamma r\pi\cos(\theta_k))e^{-\gamma r\pi\sin(\theta_k)} + e^{-2\gamma r\pi\sin(\theta_k)}}{2(1-\cos(\gamma r\pi))}} \\ &+ \mathcal{O}\left(\frac{1}{N}\right)\end{aligned} \tag{7.34}$$

from which we deduce that:

$$\begin{aligned}\frac{|\delta_k|}{|\delta_0|} &\leq e^{-rt(1-\cos(\theta_k))}\frac{1 + e^{-\gamma r\pi\sin(\theta_k)}}{\sqrt{2(1-\cos(\gamma r\pi))}} \\ &\leq \frac{e^{-rt(1-\cos(\theta_k))}}{\sqrt{1-\cos(\gamma r\pi)}}\end{aligned} \tag{7.35}$$

as $\theta_k \in [0,\pi]$. The same computation as in the case $t = o(N)$ shows that the upper bound of (7.35) is smaller than ϵ as soon as $k \leq \frac{1}{\pi}\sqrt{N\frac{2\log\epsilon\sqrt{1-\cos(\gamma r\pi)}}{\gamma r}} = o(N)$, i.e. $\theta_k = o(1)$. It is thus possible to expand the square-root term of (7.34) with respect to θ_k, and one gets (7.30) the same way as for the previous case. □

If we choose $t = N$ as suggested in [Lev81], we get $N^* = \mathcal{O}(\sqrt{N})$. The resulting algorithm reads:

Algorithm 7.15.

1. The notation $t = \Theta(N)$ means that t is bounded above and below by a linear function of N, while $t = \mathcal{O}(N)$ means that t is only bounded above by a linear function of N. Here, it is important to make this distinction as the argument in the proof is not the same if t is a $\Theta(N)$ or a $\mathcal{O}(N)$ without being a $\Theta(N)$.

Given $N \in \mathbb{N}^*$, $\boldsymbol{p} \in [0,1]^d$, $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{N}^d$ such that $\forall j \in \{1,\dots,d\}$, $0 \leq a_j \leq b_j \leq N-1$, $\epsilon_{max} > 0$, do:

1. Compute $r := \epsilon_{max}^{\frac{1}{2N_a}}$
2. Compute $\delta_0 := \prod_{j=1}^{d} \pi_{a_j b_j}^{(j)}(r) - \zeta_{N_a}^{-\sigma_a} \prod_{j=1}^{d} \pi_{a_j b_j}^{(j)}(r\zeta_{N_a})$
3. Set $v := \delta_0$, $k := 1$
4. Repeat
   (a) Compute $\delta_k := (-\zeta_{N_a}^{-\sigma_a})^k \Big(\prod_{j=1}^{d} \pi_{a_j b_j}^{(j)}(r\zeta_{N_a}^k) - \zeta_{N_a}^{-\sigma_a} \prod_{j=1}^{d} \pi_{a_j b_j}^{(j)}(r\zeta_{N_a}^{k+1}) \Big)$
   (b) Compute $v := v + \delta_k$
   (c) Set $k := k + 1$
5. Until $k = N$ or $|\delta_k| < \epsilon_{machine}|\delta_0|$
6. Return $\Re(v)/(2N_a r^N \mathbb{P}(Y_t = N))$

The evaluation of $\pi_{b_j a_j}^{(j)}$ is done using (7.13) or one of the more involved $\mathcal{O}(1)$ methods. We note that the definitions of δ_0 and δ_k is not the same as in Proposition 7.14, but using (7.23) we see that the ratio $|\delta_k|/|\delta_0|$ is the same.

It must be emphasized that in an actual implementation of this algorithm, one should include tests to detect trivial situations for which an early exit is possible. For example, when one component of $\boldsymbol{x}$ is larger than N, the problem is reduced to a lower dimensional one, or if $\sigma_b < N$ or $\sigma_a > N$, the probability is zero.

This algorithm is available in the Open TURNS software [Ope], an Open Source C++ library dedicated to probabilistic modeling and uncertainty propagation.

7.5 Numerical experiments

The objectives of these numerical experiments are to assess the accuracy of the proposed algorithm on various examples used in the literature, and to check its time complexity. All the computations have been made using the Levin recommendation for t, namely $t = N$. There is certainly more insight to be gained in the study of the influence of t on the numerical accuracy.

7.5.1 Accuracy

In this numerical experiment, we check the accuracy of the proposed algorithm on the computation of the multinomial distribution function in the following settings: $d = N$, $p_1 = \cdots = p_d = 1/d$ for $N \in \{\lfloor 2^{k/2} \rfloor \,|\, k = 2,\dots,20\}$. The distribution function is computed at the points $(x_1 = \cdots = x_d = k)$ for k such that the resulting probability value is in $[10^{-5}, 1-10^{-5}]$. For each value of N, the maximum relative error is plotted against the size N on a logarithmic scale on Figure 7.1, for ϵ_{max} taken in $\{10^{-7}, 10^{-9}, 10^{-11}\}$. The points with a zero maximal error are not plotted.

It was not possible to explore larger values for N due to the space complexity of the reference algorithm.

We note several facts from this experiment. The first one is that the accuracy of the proposed algorithm is close to the machine precision on a wide range of problem size, and even better than the accuracy of the **exact** algorithm as soon as the problem size is larger than a few tens. The second one is that the choice of ϵ_{max} does not seems to have a

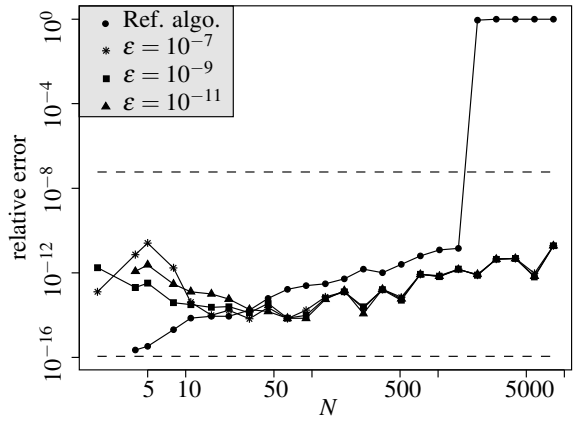


Figure 7.1: Maximum relative error for various problem size N and various precision parameter ϵ_{max}. For problem size larger than 50, the proposed algorithm is consistently more accurate than the reference algorithm, and the achieved accuracy does not depends on the value of ϵ_{max}. The two horizontal dashed lines correspond to the single and double precision accuracies.

significant impact on the accuracy of the algorithm as soon as the problem size is larger than a few tens. For smaller sizes, a value of $\epsilon_{max} \simeq 10^{-9}$ seems to give the best overall precision, even if in this case the algorithm [Fre09] should probably be preferred. The third fact is that the implementation of the algorithm [Fre09] as given in the reference paper seems to have an overflow for problems of size larger than few thousands.

7.5.2 Some classical examples

Here are the results (see Table 7.3)of our algorithm on the classical examples that can be found in [BM95], [Cor11], [Lev81] etc. The algorithm is implemented in Python, using double-precision for the computation. These results have been checked against both a Monte Carlo simulation with 10^9 samples, and the algorithm in [Fre09] using the reference implementation in R provided by the author as well as a multi-precision implementation in Maple. For each example, we give the 16 digits of the computed result and underline the digits that differ from the exact result. We also give the absolute and relative error of the computed result.

Example 7.16. This example [2] is from [BM95], which consider the classification of $N = 200$ adult subjects into $d = 4$ marital status. This example leads to the following computation:

$$\boldsymbol{X} \sim \mathcal{M}(200, [0.2, 0.35, 0.15, 0.3])$$

1. $\mathbb{P}(X_1 \leq 30, X_2 \leq 80, X_3 \leq 40, X_4 \leq 50)$

Example 7.17. This example is exposed in both [Cor11] and [Lev81], and is attributed to Mallows. It consists in the following computation:

$$\boldsymbol{X} \sim \mathcal{M}(500, [p_1 = \cdots = p_{50} = 1/50])$$

1. $\mathbb{P}(X_1 \leq 19, \ldots, X_{50} \leq 19)$

2. for which the wrong value of 0.030837 is reported (using a storage of 1373701 floating point numbers for the computation)

Example	probability	absolute error	relative error
7.16-1	$0.4784509465\underline{18295}\,10^{-5}$	$1.5\,10^{-17}$	$3.2\,10^{-12}$
	$\mathbf{0.4784509465802881\,10^{-5}}$		
7.17-1	$0.852726985258\underline{1543}$	$1.5\,10^{-14}$	$1.8\,10^{-14}$
	$\mathbf{0.8527269852581694}$		
7.17-2	$0.602684281137\underline{5376}$	$2.3\,10^{-14}$	$3.9\,10^{-14}$
	$\mathbf{0.6026842811375610}$		
7.17-3	$0.520266492592\underline{7378}$	$2.3\,10^{-14}$	$4.4\,10^{-14}$
	$\mathbf{0.5202664925927609}$		
7.18-1	$0.312632188766\underline{41}$	$1.6\,10^{-15}$	$4.9\,10^{-15}$
	$\mathbf{0.3126321887664725}$		
7.18-2	$0.837043537778\underline{8633}$	$1.0\,10^{-14}$	$1.2\,10^{-14}$
	$\mathbf{0.8370435377788733}$		

Table 7.3: Probability value and precision of the examples. The exact values, rounded to the 16th significant figure, are also given in bold face.

Based on the same multinomial distribution, these two other computations [3] are proposed in [Cor11]:

2. $\mathbb{P}\left(4 \leq X_1, \ldots, 4 \leq X_{50}\right)$
3. $\mathbb{P}\left(4 \leq X_1 \leq 19, \ldots, 4 \leq X_{50} \leq 19\right)$

Example 7.18. This example is exposed in [Lev81], and is attributed to Barton and David. It consists in the following computation:

$$\boldsymbol{X} \sim \mathcal{M}(12, [p_1 = \cdots = p_{12} = 1/12])$$

1. $\mathbb{P}\left(X_1 \leq 2, \ldots, X_{12} \leq 2\right)$
2. $\mathbb{P}\left(X_1 \leq 3, \ldots, X_{12} \leq 3\right)$

The last computation is essentially the same as the one presented in [BS98], for which the reported absolute error is of order $5\,10^{-5}$, which illustrates the limited precision of the best available approximate algorithm.

7.5.3 Time complexity assessment

The time complexity benchmark consists in the evaluation of the distribution function of the multinomial distribution in different settings for the pair (N, d). The objective is to verify the asymptotic time complexity of the algorithm with respect to both the N and d parameters in the most demanding situation, namely the computation of $\mathbb{P}\left(X_1 \leq N-1, \ldots,\ X_d \leq N-1\right)$ for equiprobable X_i.

We will test the configurations $(N, d) \in \{\lfloor 10^{k/5} \rfloor \mid k = 10, \ldots, 25\} \times \{10^k \mid k = 2, \ldots, 5\}$ for the time complexity with respect to N, and $(N, d) \in \{10^k \mid k = 2, \ldots, 5\} \times \{\lfloor 10^{k/5} \rfloor \mid k = 10, \ldots, 25\}$ for the time complexity with respect to d.

The time complexity matches perfectly the theoretical bounds, as seen on Figures 7.2 and 7.3, which confirms that the algorithm is a significant improvement over the previous ones: when $d = N$, we get a time complexity of $\mathcal{O}(N^{3/2})$ instead of $\mathcal{O}(N^3)$ for the reference algorithm.

3. for which the wrong values of resp. 0.877373 and 0.750895 are reported.

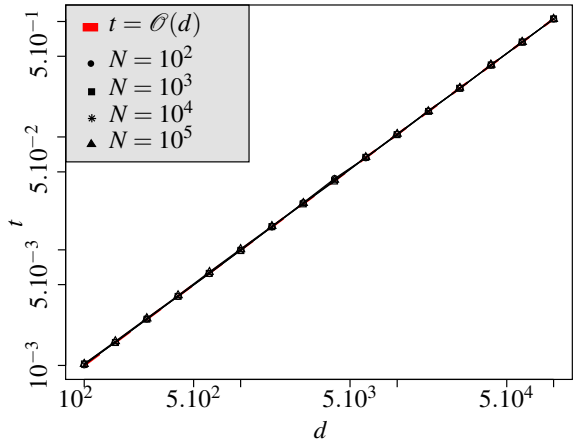


Figure 7.2: Evolution of the time complexity with respect to the dimension d in logarithmic scale, for several values of N. The time is normalized such that it is equal to 1 for the largest value of d. We see the perfect matching of the $\mathcal{O}(d)$ complexity as all the curves are superposed with the dashed line $t = d$.

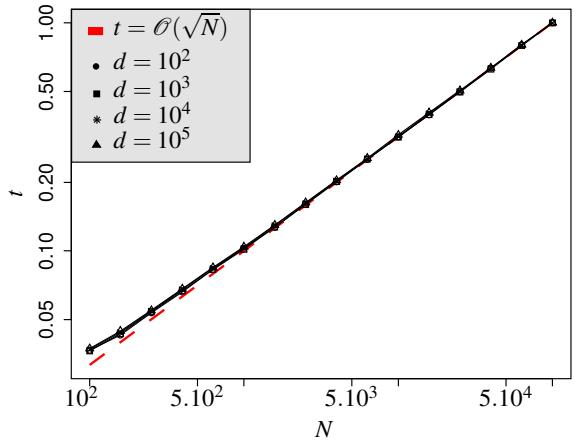


Figure 7.3: Evolution of the time complexity with respect to N in logarithmic scale, for several values of the dimension d. The time is normalized such that it is equal to 1 for the largest value of N. We see the almost perfect matching of the $\mathcal{O}(\sqrt{N})$ complexity as all the curves are almost superposed with the dashed curve $t = \sqrt{N}$.

7.6 Conclusion

In this chapter, we provide an algorithm that permits the computation of rectangular probabilities to high accuracy for a class of multi-dimensional discrete probability distributions that includes the multinomial, multi-dimensional hypergeometric and multi-dimensional Pólya distributions. This algorithm can be made exact in exact arithmetic with a constant space complexity and a polynomial time complexity that matches the best available algorithms so far.

More interestingly, its approximate version allows for significant time complexity improvement for an actual accuracy that matches and even outperforms the accuracy of previous **exact** algorithms that suffers from round-off errors when implemented in finite precision arithmetic.

Several numerical experiments have demonstrated the performances of this algorithm in the multinomial case, both with respect to its accuracy and time complexity.

This algorithm allows to address problems that were impossible to deal with using previous state-of the art algorithms, either in terms of problem size or in terms of accuracy. It has been implemented in the Open TURNS software [Ope], an Open Source software dedicated to probabilistic modeling and uncertainty propagation. It is also available as a Python [Pyt] script or as a Maple [Map] script upon request to the author.

Some additional work should be made regarding this algorithm, regarding its sensitivity to round-off error or the optimal choice for t, even if the choice $t = N$ seems to be effective in the multinomial case.

With minor modifications, this algorithm could be extended to other discrete models such as the combinatoric models considered e.g. in [ABT00].

Conclusions and perspectives

In this thesis, we have shed some insight on common practices in stochastic modeling thanks to the theory of copulas and we provided effective tools to deal with modeling situations that were previously not covered by the literature such as the modeling of dependence in presence of constraints, or only partially covered such as the efficient evaluation of the distribution function of discrete distributions in a specific class that encompasses distributions of universal use in probability and statistics.

The first work was to illustrate the necessity to fully describe the dependence structure of a random vector through its copula instead of giving only a partial description of this structure thanks to a set of measures of association. This was the main objective of chapter 2.

Using the theory of copulas, we have also highlighted the probabilistic modeling hypotheses hidden behind the common practices related to the use of iso-probabilistic transformations, namely the Nataf transformation and the Rosenblatt transformation. It has allowed us to generalize the Nataf transformation in order to relax some of these modeling hypotheses, and we have provided an extension of the Breitung asymptotic formula used in conjunction with such transformations in the context of the first or second order reliability methods. Those were the main objectives of chapters 3, 4 and 5.

We have adressed an original dependence modeling situation: the modeling of random vectors with prescribed marginal distributions and an additional constraint that must be fulfilled almost surely, in the case where this constraint is an increasing ordering between the components of the random vector. We have proved a theorem that fully describes the copulas compatible with such constraints for a given set of compatible marginal distribution functions, and we have defined a class of such copulas in the bi-dimensional case, called the subsquare copulas. We have detailed all the analytical and algorithmic aspects in order to allow the effective use of such copulas in numerical simulations. This work was the objective of chapter 6.

Finally, we have presented a new efficient and accurate algorithm to evaluate the rectangular probabilities of a class of multivariate discrete distributions that play a central role in many probabilistic and statistical situations. This algorithm allows to address high-dimensional situations essentially exactly (i.e up to machine precision) where the practitioner was previously limited to asymptotic approximations with no control on the approximation error. This was the objective of chapter 7.

For all these topics, we have identified possible extensions that we list below.

A possible extension of the work presented in chapter 2 could be an extension of Fréchet-Hoeffding bounds to the set of copulas compatible with a given dependence information, or conditions on the feasibility of a ε-synthesis with a finite dependence information for a specific class of copulas.

A possible extension of the work presented in chapters 3, 4 and 5 could be to compare the FORM and SORM approximations resulting from the use of the generalized Nataf

transformation and the Rosenblatt transformation in the case of non-Gaussian elliptical copulas. In particular, it would be interesting to see if the use of the generalized Nataf transformation provides more accurate approximations than the use of the Rosenblatt transformation, in addition to being more computationnaly efficient.

A possible extension of the work presented in chapter 6 could be to extend the results to more general constraints than an ordering constraint, for exemple a general affine inequality constraint. The construction of compatible copulas other than the sub-square or sub-hypercube copulas would be very valuable, and more generaly the geometry of the set of compatible copulas could be described in more details. Such a work has already started in collaboration with Anne Dutfoy (EDF) in the continuity of this thesis, using the sub-square copulas as building blocks for the construction of high-dimensional copulas in the spirit of vine copulas.

The ideas behind the algorithm presented in chapter 7 have a much broader scope than the one presented in this chapter. They have already been used in the definition of a highly efficient algorithm to compute essentially exactly the p-value of a new statistical test dedicated to the detection of exceptional sequences in the DNA, in collaboration with Professor Bernard Ycart (Joseph Fourier University).

Bibliography

[ABNN08] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, and H.N. Nagaraja. *A First Course in Order Statistics*. Classics in Applied Mathematics. Siam, 2008.

[ABT00] R. Arratia, A. D. Barbour, and S. Tavaré. Limits of logarithmic combinatorial structures. *The Annals of Probability*, 28(4):1620–1644, 2000.

[AGK05] J. Avérous, C. Genest, and S. C. Kochar. On the dependence structure of order statistics. *J. Multivar. Anal.*, 94(1):159–171, May 2005.

[AL82] S. F. Arnold and J. Lynch. On ali's characterization of the spherical normal distribution. *Journal of Royal Statistical Society*, 44(1):49–51, 1982.

[AW91] J. Abate and W. Whitt. The fourier-series method for inverting transforms of probability distributions, 1991.

[BB89] R. B. Bapat and M. I. Beg. Order statistics for nonidentically distributed variables and permanents. *Sankhyā:The Indian Journal of Statistics, Serie A (1961-2002*, 51(1):79–93, February 1989.

[BM95] K. J. Berry and P. W. Jr. Mielke. Exact cumulative probabilities fort the multinomial distribution. *Educational and Psychological Measurement*, 55(5):769–772, 1995.

[Bre84] K. Breitung. Asymptotic approximation for multinormal integrals. *Journal of Engineering Mechanics*, 110(3):357–366, 1984.

[Bre94] K. Breitung. *Asymptotic Approximations for Probability Integrals*. Lecture Notes in Mathematics. Springer-Verlag Berlin Inc., 1rst edition, 1994.

[BS98] R. W. Butler and R. K. Sutton. Saddlepoint approximation for multivariate cumulative distribution functions and probability computations in sampling theory and outlier testing. *Journal of the American Statistical Association*, 93(442):596–604, 1998.

[CB00] A. Childs and N. Balakrishnan. Some approximations to the multivariate hypergeometric distribution with applications to hypothesis testing. *Computational Statistics & Data Analysis*, 35(2):137–154, 2000.

[CFS07] A. Charpentier, J.-D. Fermanian, and O. Scaillet. The Estimation of Copulas: Theory and Practice. In J. Rank, editor, *Copulas: From theory to application in finance*, pages 35–62. Risk Books, 2007.

[Cor11] C. Corrado. The exact distribution of the maximum, minimum and the range of multinomial/dirichlet and multivariate hypergeometric frequencies. *Statistics and Computing*, 21:349–359, 2011.

[DKL86a] A. Der Kiureghian and Liu P. L. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2):105–112, 1986.

[DKL86b] A. Der Kiureghian and P. L. Liu. Structural reliability under incomplete probabilistic information. *Journal of Engineering Mechanics*, 112(1):85–104, 1986.

[DL09] A. Dutfoy and R. Lebrun. Practical approach to dependence modelling using copulas. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 223(4):347–361, 2009.

[DM86] A. R. DiDonato and A. H. Jr. Morris. Computation of the incomplete gamma function ratios and their inverse. *ACM Trans. Math. Softw.*, 12(4):377–393, 1986.

[DM05] O. Ditlevsen and H. O. Madsen. *Structural Reliability Methods.* John Wiley & Sons, 2005.

[DN03] H.A. David and H.N. Nagaraja. *Order Statistics.* Wiley series in probability and mathematical statistics. Probability and mathematical statistics. John Wiley, 2003.

[Dol83] K. Dolinsky. First-order second-moment approximation in reliability of structural systems : Critical review and alternative approach. *Structural Safety*, 1:211–231, 1983.

[ELM03] P. Embrechts, F. Lindskog, and A. Mcneil. *Modelling dependence with copulas and applications to risk management*, pages 329–384. North Holland, 2003.

[FJS05] G. Frahm, M. Junker, and R. Schmidt. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1):80–100, 2005.

[Fre09] J. Frey. An algorithm for computing rectangular multinomial probabilities. *Journal of Statistical Computation and Simulation*, 79(12):1483–1489, 2009.

[Gay51] A. K. Gayen. The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, 38:219–247, 1951.

[GN07] C. Genest and J. Neslehova. A primer on copulas for count data. *ASTIN Bulletin*, 37(2):475–515, 2007.

[Goo57] I. J. Good. Saddle-point methods for the multinomial distribution. *Annals of Mathematical Statistics*, 28(4):861–881, 1957.

[Hig02] N. J. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM: Society for Industrial and Applied Mathematics, 2nd edition, 2002.

[HL74] A. M. Hasofer and N. C. Lind. An exact and invariant first order reliability format. *Journal of Engineering Mechanics*, 100:111–121, 1974.

[Joe97] H. Joe. *Multivariate models and dependence concepts.* Chapman & Hall, 1997.

[Joh60] N. L. Johnson. An approximation to the multinomial distribution some properties and applications. *Biometrika*, 47(1-2):93–102, 1960.

[JR08] P. Jaworski and T. Rychlik. On distributions of order statistics for absolutely continuous copulas with applications to reliability. *Kybernetika*, 44(6):757–776, 2008.

[Kal02] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[KC06] D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling.* Wiley series in probability and statistics. John Wiley & Sons, 2006.

[KFN87] K.T. Kotz, S. Fang, and K.W. Ng. *Symmetric multivariate and related distributions*. Chapman & Hall, 1987.

[KN00] S. Kotz and S. Nadarajah. *Extrem Value Distributions Theory and Applications*. Imperial College Press, 2000.

[LD] R. Lebrun and A. Dutfoy. Which copulas for order statistics with prescribed marginal distributions? *Journal of Multivariate Analysis*, Submitted.

[LD09a] R. Lebrun and A. Dutfoy. Do rosenblatt and nataf isoprobabilistic transformations really differ? *Probabilistic Engineering Mechanics*, 24(4):577–584, 2009.

[LD09b] R. Lebrun and A. Dutfoy. A generalization of the nataf transformation to distributions with elliptical copula. *Probabilistic Engineering Mechanics*, 24:172–178, 2009.

[LD09c] R. Lebrun and A. Dutfoy. An innovating analysis of the nataf transformation from the copula viewpoint. *Probabilistic Engineering Mechanics*, 24(3):312–320, 2009.

[Leb04] R. Lebrun. Modelling dependence with copulas in reliability analysis, a new approach to the form and sorm methods. In *Proceedings of PSAM'8*. PSAM, ASME Press, June 2004.

[Leb12] R. Lebrun. Efficient time/space algorithm to compute rectangular probabilities of multinomial, multivariate hypergeometric and multivariate pólya distributions. *Statistics and Computing*, 2012.

[Lem05] M. Lemaire. *Fiabilité des structures*. Hermes, 2005.

[Lev81] B. Levin. A representation for multinomial cumulative distribution functions. *Annals of Statistics*, 9(5):1123–1126, 1981.

[Lev83] B. Levin. On calculations involving the maximum cell frequency. *Communications in Statistics - Theory and Methods*, 12(11):1299–1327, 1983.

[Lev92] B. Levin. In re: "siobhan's problem: The coupon collector revisited". *The American Statistician*, 46(1):76, 1992.

[Lev06] H. Levy. *Stochastic Dominance: Investment Decision Making Under Uncertainty*. Studies in Risk And Uncertainty. Springer, 2006.

[Ma99] C. Ma. Representations, bounds and approximations for tail probabilities of multivariate non-central hypergeometric and negative hypergeometric distributions. *Journal of Statistical Computation and Simulation*, 62:237–258, 1999.

[Mal68] C. L. Mallows. An inequality involving multinomial probabilities. *Biometrika*, 55(2):422–424, 1968.

[Map] Maplesoft. Maple computer algebra system.

[Nat62] A. Nataf. Détermination des distributions de probabilités dont les marges sont données. *Comptes Rendus de l'Académie des Sciences*, A 225:42–43, 1962.

[Nel06] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer-Verlag New York Inc., 2nd edition, 2006.

[NS10] J. Navarro and F. Spizzichino. On the relationships between copulas of order statistics and marginal distributions. *Statistics & Probability Letters*, 80(5–6):473–479, 2010.

[Ope] OpenTURNS. Open source plateform for the treatment of uncertainties, risk's and statistics.

[Pyt] Software Foundation Python. Python programming language.

[Ros52] M. Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, 1952.

[RT94] L. Rüschendorf and W. Thomsen. Note on the schrödinger equation and I-projections. *Statistics and Probability Letters*, 17:369–375, 1994.

[RTC96] Committee SC-117 RTCA. Portable electronic devices carried onboard aircraft. Technical Report DO-233, RTCA, August 1996.

[Rüs09] L. Rüschendorf. On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139:3921–3927, 2009.

[Sca84] M. Scarsini. On measures of concordance. *Stochastica*, 8:201–218, 1984.

[Skl59] M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publication de l'Institut Statistique Universitaire Paris*, 8:229–231, 1959.

[SS04] R. Schmidt and U. Stadtmüller. Nonparametric estimation of tail dependence. Technical report, University of Cologne, 2004.

[Str65] V. Strassen. Existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36:423–439, 1965.

[SW81] B. Schweizer and E.F. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879–885, July 1981.

[Tem94] N. M. Temme. A set of algorithms for the incomplete gamma functions. *Probability in the Engineering and Informational Sciences*, 8:291–307, 1994.

[Tve88] L. Tvedt. Second order reliability by an exact integral. In *Proceedings of the IFIP Working Conference reliability and optimization of Structural Systems*, pages 377–384. Thoft-Christensen, 1988.

[Tve90] L. Tvedt. Distribution of quadratic forms in normal spaces - application to structural reliability. *Structural Safety*, 3:143–166, 1990.

Index